POLITECNICO DI TORINO
Repository ISTITUZIONALE

Mid-term Performance and Calibration of Multi-sensor Low-Cost Systems for Air Quality Monitoring in Urban Environments

Original
Mid-term Performance and Calibration of Multi-sensor Low-Cost Systems for Air Quality Monitoring in Urban Environments / Fellini, S., Gallione, D., Vaccaro, V., Mastromatteo, N., Fattoruso, G., Clerico, M., Salizzoni, P.. - In: AEROSOL AND AIR QUALITY RESEARCH. - ISSN 1680-8584. - ELETTRONICO. - 26:(2026). [10.1007/s44408-026-00150-1]

Availability:
This version is available at: 11583/3013519 since: 2026-07-24T12:48:07Z

Publisher:
Springer Nature

Published
DOI:10.1007/s44408-026-00150-1

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

25 October 2026

Aerosol and Air Quality Research (2026) 26:73
https://doi.org/10.1007/s44408-026-00150-1

RESEARCH

Mid-term Performance and Calibration of Multi-sensor Low-Cost Systems for Air Quality Monitoring in Urban Environments

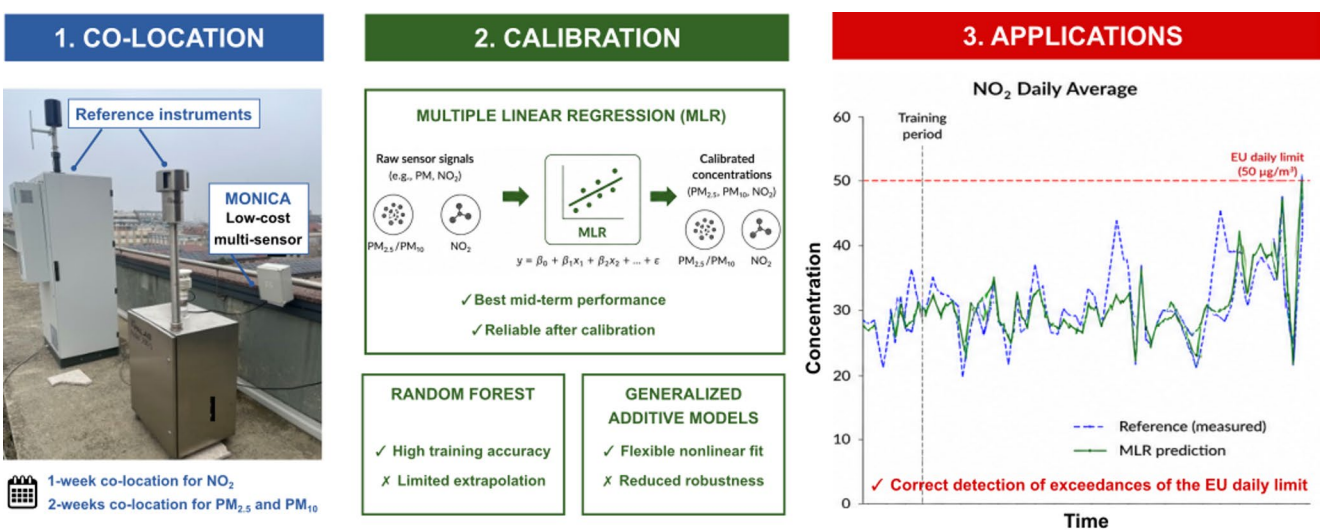
Sofia Fellini[1] · Davide Gallione[1] · Vincenzo Vaccaro[1] · Nicole Mastromatteo[1] · Grazia Fattoruso[3] · Marina Clerico[1] et al. *[full author details at the end of the article]*

Received: 21 July 2025 / Revised: 2 May 2026 / Accepted: 23 June 2026 / Published online: 3 July 2026
© The Author(s) 2026

Abstract

This study evaluates mid-term calibration strategies for MONICA, a compact low-cost multi-sensor device for urban air quality monitoring, in the context of the upcoming EU Air Quality Directive 2024/2881. A key objective is to optimize the trade-off between calibration duration and long-term monitoring performance, balancing statistical robustness with practical deployment constraints. This question is particularly relevant in mid-latitude regions, where seasonal variability may introduce biases if calibration and observation periods are misaligned. Over a three-and-a-half-month winter co-location with reference-grade instruments, we assessed three models—Multiple Linear Regression (MLR), Random Forest (RF), and Generalized Additive Models (GAM)—across three pollutants: $PM_{2.5}$, PM_{10}, and NO_2. MLR emerged as the most stable model for temporal extrapolation, while RF and GAM, although accurate short-term, showed performance degradation outside the training range. A two-week calibration period was sufficient for PM, whereas NO_2 required only one week. Although sensor accuracy declines over time-especially for NO_2—the MONICA system remains effective in tracking temporal trends and identifying regulatory exceedances. These results support the development of efficient and scalable low-cost sensor networks, offering practical insights for planning reliable air quality monitoring campaigns.

Graphical abstract



Keywords Low-cost sensors · Air quality monitoring · Sensors calibration · Machine learning regression · Sensor co-location

Sofia Fellini and Davide Gallione have contributed equally to this work.

1 Introduction

Air pollution is a leading environmental health risk, contributing to millions of premature deaths annually through respiratory and cardiovascular diseases e.g., (Di et al. 2017; Hystad et al. 2020; Lin et al. 2021; De Bont et al. 2024). Urban areas are particularly vulnerable due to high emission densities and complex morphologies that limit pollutant dispersion (Salizzoni et al. 2009; Fellini et al. 2021).

Among urban air pollutants, particulate matter (PM) and nitrogen dioxide (NO_2) remain the most critical due to their documented health effects and widespread exceedance of regulatory thresholds (Kim et al. 2015; Chen and Hoek 2024; Kang and Choi 2024). $PM_{2.5}$ is of particular concern due to its ability to penetrate deep into the respiratory system (Pope and Dockery 2006; WHO 2021), while NO_2 is strongly associated with respiratory inflammation (Khreis et al. 2017; Anenberg et al. 2022; Soulhac et al. 2023). These pollutants are regulated under the updated European Directive 2024/2881 (European Parliament and Council 2024), which tightens air quality standards in line with WHO recommendations and the European Green Deal's zero-pollution ambition.

Accurately estimating ground-level concentrations in urban areas is essential for assessing individual exposure. However, pollutants exhibit high spatiotemporal variability (Steinle et al. 2015; Fellini et al. 2024) that traditional, sparse monitoring networks, managed by environmental agencies, cannot fully capture. Consequently, understanding local air quality often relies on modeling approaches, including deterministic (McHugh et al. 1997; Soulhac et al. 2011; Kim et al. 2018; Suter et al. 2021) and statistical models (Hoek et al. 2008; Beelen et al. 2013; Ma et al. 2024), whose application is often constrained by the available input data and the required spatial resolution.

In recent years, however, the emergence of low-cost sensors (LCS) has enabled high-resolution air quality monitoring, particularly in complex and heterogeneous urban environments (Liu et al. 2024). Characterized by their compactness, low energy requirements, and affordability (Schmitz et al. 2024), LCS are increasingly used to create large-scale real-time monitoring networks that integrate both indoor and outdoor air quality measurements (Tang et al. 2024). LCS systems have been successfully used to monitor air quality in areas lacking official stations, demonstrating their potential to enhance exposure assessments across diverse environments (Cross et al. 2017; Thomas et al. 2018; Johnson et al. 2018; Hagan et al. 2018).

Despite their advantages, low-cost sensors still face significant challenges related to accuracy, precision, and durability (Badura et al. 2018; Cavaliere et al. 2018; Wang et al. 2015; Bulot et al. 2019; Giordano et al. 2021). To address these, LCS are commonly co-located with reference instruments for a calibration period, during which intra-sensor consistency (repeatability over time) and inter-sensor consistency (agreement between sensors and references) are assessed (Tang et al. 2024). Importantly, the selection of a suitable calibration model is a critical step, and its appropriateness depends on the specific pollutant being measured.

PM sensor calibration commonly employs multiple linear regression and machine learning techniques (Venkatraman Jagatha et al. 2021; Zimmerman 2022; Tang et al. 2024). These models often incorporate environmental variables like temperature and humidity, which are known to significantly influence sensor output (Crilley et al. 2018; Di Antonio et al. 2018; Macías-Hernández et al. 2023). Research includes long-term co-location campaigns using large sensor datasets (Datta et al. 2020; Giordano et al. 2021) and assessment of the impact of specific conditions such as particle sources or relative humidity on measurement performance (Badura et al. 2018; Bulot et al. 2019). Reviews by Venkatraman Jagatha et al. (2021) and Giordano et al. (2021) have highlighted both the promise and variability in sensor behavior and calibration strategies.

For NO_2, the literature is more limited and the calibration performances are generally less robust. Previous studies have employed multiple linear regression (Jiao et al. 2016; Mijling et al. 2018; Ratingen et al. 2021), while others have tested machine learning approaches, including random forests (Nowack et al. 2021; Kim et al. 2022) and neural networks (Malings et al. 2019). These works underline the importance of both long-term calibration and the inclusion of environmental covariates to mitigate sensor drift and improve prediction accuracy.

Although previous studies have provided important insights, several key limitations remain. In particular, few works have systematically evaluated calibration strategies for integrated multi-sensor systems measuring both PM and gaseous pollutants simultaneously (Jiao et al. 2016; Cross et al. 2017; Gäbel et al. 2022; Cavaliere et al. 2023). This is a significant gap, given the increasing use of such platforms in urban air quality networks. Moreover, calibration duration is seldom systematically assessed, despite its crucial role in ensuring model reliability under the temporal constraints typical of open-field monitoring deployments. There is currently no standardized co-location duration, and the reported co-location durations in recent works have varied from several days to several months (Levy Zamora et al. 2023). This issue is especially relevant for non-linear and data-driven models which may suffer from reduced generalizability when applied beyond the temporal range of their training data (Nowack et al. 2021). The long-term stability of calibrated sensors and the performance decay also remain poorly understood. These uncertainties hinder the effective

deployment of LCS, especially in long-term monitoring where recalibration is costly or logistically challenging.

To address these limitations, this study evaluates the MONICA device (De Vito et al. 2020, 2021; Fattoruso et al. 2022; De Vito et al. 2023; Ramel-Delobel et al. 2024). This integrated platform allows for multi-pollutant monitoring, relevant for urban-scale networks. We assess three calibration models (Multiple Linear Regression, MLR; Random Forest, RF; and Generalized Additive Models, GAM) across NO_2, $PM_{2.5}$, and PM_{10}. The analysis addresses three core questions: (1) which calibration models provide the highest accuracy for each pollutant; (2) what is the minimum co-location period required to achieve stable and reliable calibration, depending on the model; and (3) for how long calibrated sensors maintain acceptable performance over time. A key focus of the study is to determine the maximum practical ratio between the calibration phase (2) and the total observation period (3). This is especially relevant for two reasons. First, in mid-latitude regions, extending the calibration period over several weeks or months may introduce seasonal biases if environmental conditions differ between the training and deployment phases. Second, given the limited lifespan of LCS (typically 1–2 years), an excessively long co-location phase limits the device's actual monitoring utility.

The campaign was conducted in the city of Turin (Po Valley), one of Europe's major air pollution hotspots, where persistent pollution levels are driven by intense local emissions and unfavorable meteorological conditions (Bo et al. 2020).

This work offers operational guidance for LCS-based campaigns, optimizing calibration strategies for robust and scalable monitoring networks for urban air pollution assessment.

2 Material and Methods

The experimental campaign was conducted in Turin, Italy, over a period of approximately four months, from November 6, 2023, to February 25, 2024. During this time, the MONICA multi-sensor systems were co-located with certified reference instruments at the DIATI (Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino) monitoring station. This period allowed the systems to be exposed to a wide range of typical urban winter concentrations: hourly averaged concentrations of NO_2 ranged from 3 to 85 μg/m^3, while PM_{10} and $PM_{2.5}$ varied between 2–132 μg/m^3 and 1–132 μg/m^3, respectively. These data provided the basis for both the development (training) of the regression models and the subsequent validation of their temporal stability, as the sensors remained in the same location throughout the entire campaign.

2.1 MONICA Low Cost Sensors

The ENEA MONICA device is a compact, cost-effective microsensor designed for air quality monitoring in both fixed and mobile applications. It is housed in a robust, waterproof enclosure (see box in Fig. 1) with mesh-protected inlets to allow air exchange while preventing contamination from insects or debris. Inside the box, the monitoring setup includes both the MONICA sensor and a Raspberry Pi unit. The Raspberry Pi communicates with the MONICA device via Bluetooth to receive raw data, and–when connected to the local Wi-Fi network–it transfers the collected data to a cloud-based storage and visualization platform. The Raspberry Pi also allows for local access to raw files and remote control via RealVNC, enabling troubleshooting and system monitoring when necessary.

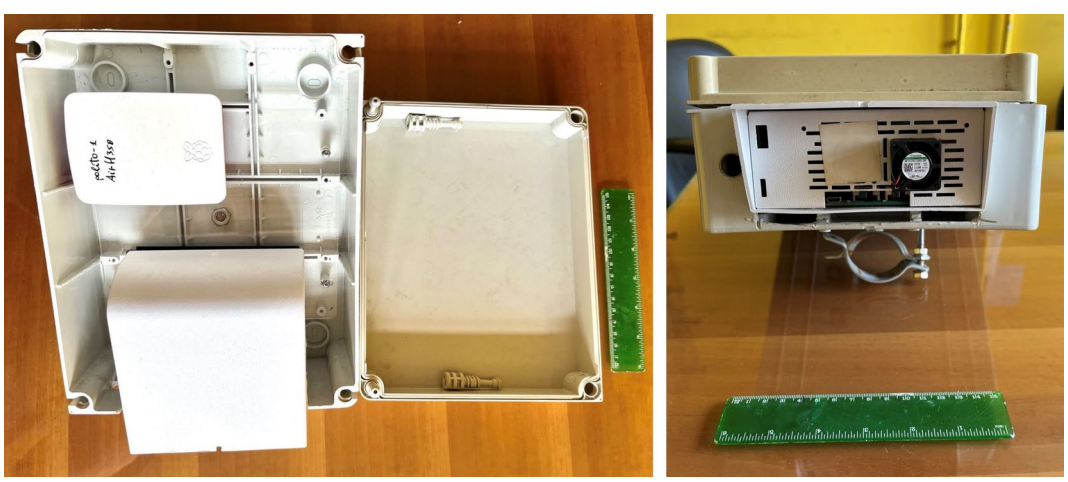


Fig. 1 ENEA MONICA device. Left: Internal view of the device. Right: Close-up of the chamber housing the PM and NO_2 microsensors. The overall dimensions of the device are 19 × 24 × 10 cm^3, and the microsensor chamber measures 12.5 × 13.5 × 6 cm^3. A 15 cm ruler is placed next to the device for scale reference

The MONICA system incorporates a Plantower PMS7003 optical particle counter (Beijing Plantower Co., Ltd 2024) and several electrochemical sensors for gas monitoring: the Alphasense CO-A4 (Alphasense - AMETEK Inc. 2024) for carbon monoxide, the NO2-A43F (Alphasense - AMETEK Inc. 2024) for nitrogen dioxide, and the OX-A431 (Alphasense - AMETEK Inc. 2026) for ozone detection. The MONICA device also records meteorological parameters such as temperature (T) and relative humidity (rH) via a DHT22 sensor (Guangzhou Aosong Electronics Co., Ltd. 2026). Details regarding the measurement ranges for the sensors deployed in this study are summarized in Table 1.

The MONICA system estimates NO_2 concentrations using the voltage signals from two electrochemical sensors: the NO_2-selective sensor (NO2-A43F) and the OX-A431 sensor, which is primarily sensitive to ozone but also exhibits cross-sensitivity to NO_2. Both sensors (Alphasense A4 series) provide outputs from a working electrode (WE) and an auxiliary electrode (AE); the latter is specifically designed to physically compensate for temperature-dependent baseline shifts (Spinelle et al. 2015; Castell et al. 2017). The corresponding signals (NO_2WE, NO_2AE, NO_2O_3WE, NO_2O_3AE) are used to correct the NO_2 response by accounting for both thermal drift and ozone interference (Spinelle et al. 2015; Castell et al. 2017).

Particulate matter concentrations are derived from the optical particle counter. It provides mass concentrations under standard conditions ($PM_{2.5}$SP, PM_{10}SP in Table 1), assuming a reference particle size distribution and dry aerosol state. Since field conditions rarely conform to these assumptions, the system also provides adjusted values ($PM_{2.5}$AE, PM_{10}AE in Table 1) that account for ambient temperature and relative humidity. These corrections are applied using a factory-defined regression model, which internally adjusts the raw signals based on typical atmospheric responses. However, this model is not site-specific: the regression parameters are predefined by the manufacturer and remain fixed across locations and conditions.

Additionally, the MONICA device reports particle number concentrations for particles larger than 2.5 µm ($NP_{2.5}$) and 10 µm (NP_{10}), which provide complementary information on aerosol number and size distribution.

Data collected by MONICA are transmitted in real-time to a dedicated web platform, where users can visualize measurements through time-series graphs and location-based air quality maps.

2.2 Certified Instruments for Sensor Calibration and Validation

Two air quality monitoring stations were used to provide reference measurements for the calibration and validation of the MONICA sensors: the fixed station CC-Green Roof Lab (Fig. 2, left) and the mobile laboratory CC-TrAIRer Lab (Boanini et al. 2021) (Fig. 2, right). Both stations are equipped with certified instruments for air quality and meteorological monitoring.

Specifically, the following reference instruments were used in this study: a Palas Fidas 200 S (2 in Fig. 2) for particulate matter ($PM_{2.5}$ and PM_{10}), an Envea AC32e analyzer (3 in Fig. 2) and a Serinus 40 analyzer (4 in Fig. 2) for NO_2 , and a Davis Vantage Pro 2 weather station (5 in Fig. 2) for temperature, relative humidity, precipitation, wind speed, and direction.

The Palas Fidas 200 S is an optical spectrometer designed for atmospheric pollution monitoring, certified according to the EN 16450 standard. It enables continuous and simultaneous measurement of various particulate fractions: PM_1, $PM_{2.5}$, PM_4, PM_{10}, and total suspended particles (TSP), along with particle number concentration and size

Table 1 Sensor output parameters from the ENEA MONICA low-cost system relevant to the measurement of NO $_2$, $PM_{2.5}$, PM_{10}, temperature and humidity

Nitrogen dioxide – NO_2		
NO_2WE	Terminal voltage (mV)—Working Electrode, sensitive primarily to NO_2	Alphasense NO2-A43F Range: 0–20 ppm
NO_2AE	Terminal voltage (mV)—Auxiliary Electrode, for interference compensation	
NO_2O_3WE	Terminal voltage (mV)—Working Electrode, sensitive to O_3 and NO_2	Alphasense OX-A431 Range: 0–20 ppm
NO_2O_3AE	Terminal voltage (mV)—Auxiliary Electrode, for interference compensation	
Particulate Matter—PM		
$PM_{2.5}$SP	$PM_{2.5}$ concentration (µg/m^3)—under standard particle conditions	Plantower PMS7003 Range: 0–500 µg/m^3
PM_{10}SP	PM_{10} concentration (µg/m^3)—under standard particle conditions	
$PM_{2.5}$AE	$PM_{2.5}$ concentration (µg/m^3) under atmospheric conditions	
PM_{10}AE	PM_{10} concentration (µg/m^3) under atmospheric conditions	
$NP_{2.5}$	Number of particles with diameter > 2.5 µm	
NP_{10}	Number of particles with diameter > 10 µm	
Meteorological Parameters		
rH	Relative Humidity (%)	DHT22. Range: 0–100% RH; [−40, +80] °C
T	Air Temperature (°C)	

Sensor models and their manufacturer-specified measurement ranges are provided in the final column

Fig. 2 Positioning of ENEA MONICA microsensors and reference instruments for calibration at DIATI—Politecnico di Torino. Left: CC-Green Roof. Right: CC-TrAIRer Lab. 1—ENEA MONICA™ microsensors, 2—Palas Fidas 200 S optical sensor for PM measurement, 3—Sampling head for ENVEA AC32e analyzer for NO_2 measurements, 4—Sampling head for Serinus 40 analyzer for NO_2 measurements, 5—Davis Vantage Pro 2 weather station

distribution. The measurement range is 0.18 µm to 18 µm. For this study, only $PM_{2.5}$ and PM_{10} data were used.

The Serinus 40 NOx Analyzer is a high-precision instrument designed to measure nitrogen oxides using the gas-phase chemiluminescence method. Ambient air, pre-filtered through 47 mm Teflon filters to remove large particles, enters the reaction cell through two alternating pathways dedicated to nitric oxide (NO) and nitrogen oxides (NO_x) measurements. In the first pathway, atmospheric NO reacts with O_3, producing a chemiluminescent signal proportional to its concentration. In the second pathway, the airflow passes through a delay loop and an NO_2-to-NO converter. The delay loop, a coiled flow circuit, ensures proper separation of the two measurement sequences, enabling the sequential detection of NO and total NO_x.

The ENVEA AC32e is an eco-designed reference air quality monitor that utilizes the chemiluminescence principle, which is the standard method for measuring NO_2 and NO concentrations in ambient air (certified according to EN 14211). To measure NO_x, the instrument passes the sample through a heated catalyst, which converts all nitrogen oxides into NO. By alternately introducing and bypassing the catalyst in the sample path, the analyzer compares the resulting signals, allowing for the indirect quantification of NO_2.

The Davis Vantage Pro 2 weather station is designed to monitor and record data related to the weather and climate conditions of a location, with variable frequency. The measured parameters include temperature, atmospheric pressure, humidity, wind speed and direction, solar radiation, rainfall rate, and ultraviolet radiation.

2.3 Regression Models for Calibration of the LCS

Accurate calibration of the MONICA low-cost sensor system is essential to correct for environmental interferences and sensor-specific variability (Spinelle et al. 2015; Castell et al. 2017). In this study, the MONICA device was colocated with the DIATI reference stations described in Sect. 2.2. During the co-location period, MONICA sensor outputs were paired with hourly reference measurements of NO_2, $PM_{2.5}$, and PM_{10}. These data were used to train regression models capable of translating raw sensor signals (e.g., electrode voltages and particle counts) into calibrated pollutant concentrations. This in-field calibration addresses limitations of factory-set models, which do not adapt to local environmental conditions (Lewis et al. 2018; Levy Zamora et al. 2023).

We applied three different regression models commonly used in the literature to calibrate the MONICA devices.

The first model is a Multiple Linear Regression (MLR) (Neter et al. 1996) of the following form:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n, \quad (1)$$

where Y is the predicted target variable, α_0 is the intercept, $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the coefficients for each predictor, and $X_1, X_2, \ldots, X_n$ are the input features.

The second model is a Generalized Additive Model (GAM) with smoothing splines applied to each predictor (Wood 2017). GAMs extend linear models by allowing non-linear relationships, providing a more flexible framework to capture subtle trends in the data. The equation for a GAM can be represented as:

$$Y = f_1(X_1) + f_2(X_2) + \cdots + f_n(X_n), \quad (2)$$

where Y is the predicted target variable, $f_1(X_1), f_2(X_2), \ldots, f_n(X_n)$ are smoothing functions of each predictor.

The third model is a Random Forest (RF) regressor. Random forests are ensemble methods that combine multiple decision trees to improve predictive accuracy and mitigate overfitting through averaging (Breiman 2001a). The general equation is:

$$Y = \frac{1}{n}\sum_{i=1}^{n} T_i(X), \tag{3}$$

where Y is the predicted target variable, n is the number of trees, $T_i(X)$ is the output of the ith decision tree, and X is the input feature vector. This approach is particularly effective for capturing non-linear dependencies and complex interactions between variables. The RF model was implemented using the *scikit-learn* library in Python. The number of trees (200) was selected based on a sensitivity analysis where the Root Mean Square Error (RMSE) was evaluated as a function of the number of trees. We observed that the model performance stabilized well before reaching 200 trees; therefore, this value was chosen as an optimal trade-off between predictive accuracy and computational efficiency. While increasing the number of trees typically does not, by itself, induce overfitting (as predictions are obtained through ensemble averaging (Breiman 2001b)) it is recognized that beyond a certain complexity, adding further trees generally yields diminishing returns in performance while increasing the computational burden. Consequently, the chosen threshold was deemed sufficient to ensure both model stability and generalizability without unnecessary complexity. To ensure reproducibility, a fixed random seed (*random_state* = 42) was applied, while other hyperparameters were kept at their default settings to maintain a robust configuration for capturing non-linear relationships.

The calibration was performed using continuous time windows rather than shuffled data to prevent temporal data leakage and to evaluate the sensors' performance under realistic deployment conditions (Nowack et al. 2021). For the same reason, the "time of day" was not included as a predictor; avoiding site-specific temporal features reduces the risk of overfitting to local emission cycles and ensures better model transferability across different locations and periods.

For each pollutant, the regression models initially incorporated all meteorological parameters (temperature and relative humidity) along with pollutant-related parameters from the MONICA system (Table 1).

To optimize the model and ensure its robustness, a systematic feature selection was performed using an all-subset regression approach (implemented via the `bestlm` R function). This procedure evaluated all possible combinations of predictors to identify the most effective subset for each pollutant. The selection process was based on maximizing the model's predictive accuracy while maintaining parsimony; specifically, parameters that did not provide a statistically significant improvement to the regression performance were subsequently excluded.

For illustrative purposes, the following multiple linear regression (MLR) equations correspond to the model calibrated using a two-week training period, which–as will be discussed in the following sections–represents a good trade-off between calibration effort and predictive performance:

$$\mathrm{PM}_{2.5} = \alpha_0 + \alpha_1 \cdot \mathrm{PM}_{2.5}\mathrm{AE} + \alpha_2 \cdot \mathrm{NP}_{2.5} + \alpha_3 \cdot \mathrm{rH} \tag{4a}$$

$$\mathrm{PM}_{10} = \beta_0 + \beta_1 \cdot \mathrm{PM}_{10}\mathrm{AE} + \beta_2 \cdot \mathrm{PM}_{10}\mathrm{SP} + \beta_3 \cdot \mathrm{rH} \tag{4b}$$

$$\mathrm{NO}_2 = \gamma_0 + \gamma_1 \cdot \mathrm{NO}_2\mathrm{WE} + \gamma_2 \cdot \mathrm{NO}_2\mathrm{O}_3\mathrm{WE} + \gamma_3 \cdot \mathrm{rH}. \tag{4c}$$

The regression coefficients from the two-week training period are reported in Table 2. These are dimensional coefficients, as the regression models were trained using the predictors in their original units without normalization.

The same predictor variables shown in Eqs. 4a-c were also used in the generalized additive model (GAM) and the random forest (RF) model.

As shown in Eqs. 4a-c, and in line with previous studies (Ramel-Delobel et al. 2024), relative humidity emerges as a key factor influencing sensor response. In contrast, no explicit dependence is observed between pollutant concentrations and temperature. This outcome, resulting from the feature selection process, can be explained by the specific conditions of the winter campaign, where the observed temperature range was relatively narrow (0.6 to 20.9 °C), and a moderate Pearson correlation coefficient of 0.53 was found between T and rH. Under these specific conditions, the variance associated with temperature may be sufficiently captured by the humidity parameter, rendering T a redundant predictor for the algorithms. Furthermore, as

Table 2 Coefficients for MLR model (Eqs. 4) obtained from the two-week training period

Pollutant	Regression coefficients			
$PM_{2.5}$	$\alpha_0 = 8.420$	$\alpha_1 = 0.516$	$\alpha_2 = -0.036$	$\alpha_3 = -0.090$
PM_{10}	$\beta_0 = 16.207$	$\beta_1 = 0.541$	$\beta_2 = -0.039$	$\beta_3 = -0.152$
NO_2	$\gamma_0 = -1328.265$	$\gamma_1 = 4.468$	$\gamma_2 = -0.137$	–

The coefficients are dimensional and their magnitude reflects the specific scales of the raw sensor outputs (e.g., mV or particle counts)

detailed in Sect. 2.1, the MONICA sensors already incorporate hardware and firmware-level thermal compensations. Among the pollutant-related parameters, some were found to be more influential than others. For instance, the $NP_{2.5}$ parameter, which represents the number of $PM_{2.5}$ particles, was included in the regression model for $PM_{2.5}$ concentration. In contrast, the corresponding NP_{10} parameter for PM_{10} was excluded, as it did not contribute to improving the model's accuracy.

As shown in Eq. 4c and Table 2, the NO_2 model explicitly incorporates the NO_2O_3WE signal from the OX-A431 sensor to compensate for ozone interference. While O_3 levels were typically low during this winter campaign, including these terms ensures that the model is chemically consistent and remains robust against localized or transient fluctuations in ozone concentrations. Nevertheless, further testing during summer months would be necessary to fully assess the sensors' performance under higher O_3 concentrations and broader temperature fluctuations, which could alter the weight of the ozone-correction term. In such scenarios, the inclusion of T might also become essential to maintain model stability.

2.4 Data Processing and Performance Metrics

Although the monitoring campaign spanned the entire period from November 2023 to February 2024 (as described at the beginning of this section), several interruptions affected the operation of the reference instruments, reducing the time windows available for direct comparisons. Specifically, for $PM_{2.5}$ and PM_{10}, comparisons were possible from 6 November 2023 to 5 January 2024, and again from 25 January to 25 February 2024. For NO_2, valid comparison data were available from 6 November 2023 to 5 January 2024, and from 25 January to 4 February 2024.

For the training and evaluation periods, the analysis was organized by full weeks (Monday to Sunday), as field measurement campaigns are typically structured in this manner. As shown in Table 5 in the A, there were 12 full weeks available for particulate matter ($PM_{2.5}$ and PM_{10}) and 9 full weeks for NO_2. Only the weeks with complete data coverage are numbered.

Table 3 Mathematical definitions of the performance metrics used for sensor validation

Metric	Definition
R^2	$1 - \frac{\sum_{i=1}^{N}(C_i - C_{\mathrm{ref},i})^2}{\sum_{i=1}^{N}(C_i - \overline{C})^2}$
RMSE	$\left[\frac{1}{N}\sum_{i=1}^{N}(C_i - C_{\mathrm{ref},i})^2\right]^{0.5}$
FB	$\frac{\sum_{i=1}^{N}\left(\overline{C_i} - \overline{C}_{\mathrm{ref},i}\right)}{\sum_{i=1}^{N}\overline{C}_{\mathrm{ref},i}}$
NFB	$\frac{2\sum_{i=1}^{N}(C_i - C_{\mathrm{ref},i})}{\sum_{i=1}^{N}(C_i + C_{\mathrm{ref},i})}$
MGB	$\exp\left(\frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{C_i}{C_{\mathrm{ref},i}}\right)\right)$

To assess the agreement between measurements from the MONICA sensors and the reference instruments, we employed standard statistical indicators commonly recommended in the literature for evaluating correlation, error, and bias (Kang and Choi 2024). These include the coefficient of determination (R^2), the Fractional Bias (FB), the Root Mean Square Error (RMSE), the Normalized Fractional Bias (NFB), and the Mean Geometric Bias (MGB). The selection of these metrics allows for a comprehensive evaluation of the sensor's accuracy, precision, and systematic errors, which are essential for low-cost sensor validation.

The definitions and mathematical expressions for these metrics are presented in Table 3, where the following symbols are used in the definitions: C_i denotes the concentration measured by the low-cost sensor for the i-th observation, $C_{\mathrm{ref},i}$ denotes the concentration measured by the reference instrument for the i-th observation, $\overline{C}$ is the mean concentration provided by the low-cost sensor, $\overline{C_{\mathrm{ref}}}$ is the mean concentration provided by the reference instrument, and N represents the number of paired measurements.

R^2 quantifies the extent to which sensor measurements explain the variance in reference data and is widely used to evaluate correlation strength. Values above 0.7 indicate strong agreement, while values below 0.5 reflect weak predictive performance (Moore and Kirkland 2007). The Root Mean Square Error (RMSE) measures the average magnitude of error between sensor and reference measurements. As a key indicator of overall accuracy, lower RMSE values denote better agreement (Holstius et al. 2014; Mei et al. 2020). In addition, the FB and NFB are commonly used indicators in sensor validation studies e.g., (Lopez-Restrepo et al. 2021; Bisignano et al. 2022) to assess systematic errors between sensor and reference measurements. FB indicates relative bias, helping to detect consistent over- or underestimation, while NFB normalizes this measure, allowing for comparisons across sensors and varying concentration levels. Finally, we consider the geometric mean of the ratio between sensor and reference measurements (MGB). Unlike additive bias metrics, MGB is particularly valuable for capturing multiplicative errors, common in environmental measurements, where inaccuracies often scale with concentration levels. This is especially relevant when pollutant concentrations span multiple orders of magnitude, as traditional error metrics may disproportionately emphasize high-concentration deviations. By providing a scale-invariant measure of the bias, MGB ensures a more

balanced evaluation across different concentration ranges (Chang and Hanna 2004).

3 Results

This section begins by describing the meteorological and air quality conditions observed during the measurement campaign (Sect. 3.1). It then evaluates the performance of different regression models applied to the MONICA system data (Sect. 3.2), followed by an analysis of the temporal stability of the calibration and the minimum duration required to achieve effective calibration (Sect. 3.3). Finally, it examines the capability of the low-cost sensors to detect exceedances of air quality thresholds defined by the new European Directive 2024/2881 (Sect. 3.4).

3.1 Temporal Analysis of Meteorological Data and Pollutant Concentration

Figure 3 illustrates the meteorological and air quality trends measured by the reference instruments (Sect. 2.2) during the monitoring period. For better visual clarity and to ensure a continuous time series, the figure focuses on the measurements collected from early November to late December 2023.

As is typical for the winter season in northern Italy, temperatures ranged from 0 to 15 °C during the day, with occasional peaks nearing 20 °C and an average temperature of approximately 8 °C. The lowest temperatures were observed in the first week of December. Relative humidity ranged between 14% and 95%, maintaining above 60% in the first half of November and during the central part of December. Rainfall was almost absent.

Pollutant concentrations reveal significant temporal variability over the monitoring period, with particularly elevated levels of PM and NO_2 during the colder months. In early November, $PM_{2.5}$ reached daily maximum values around 30 μg/m^3, and PM_{10} peaked at approximately 50 μg/m^3. In the second half of November, more pronounced fluctuations were observed, with daily concentrations reaching 50 μg/m^3 for $PM_{2.5}$ and 75 μg/m^3 for PM_{10}. A further rise occurred in late December, where $PM_{2.5}$ remained consistently above 40 μg/m^3 and peaked at 80 μg/m^3, while PM_{10} exceeded 50 μg/m^3 on most days and reached values up to 100 μg/m^3. However, considering the entire three-and-a-half-month campaign, pollution events were even more acute, with hourly peaks reaching 132 μg/m^3 for $PM_{2.5}$ and 132 μg/m^3 for PM_{10}.

NO_2 concentrations followed a similar trend, gradually increasing through November and reaching hourly peaks of 50 μg/m^3. In late December, NO_2 levels rose further, with hourly maximum values approaching 80 μg/m^3. The absolute hourly maximum for the full campaign was recorded at 85 μg/m^3.

The green dashed lines in the figure represent the daily mean limit values established by the new European Directive (EU) 2024/2881. During the monitoring period, the

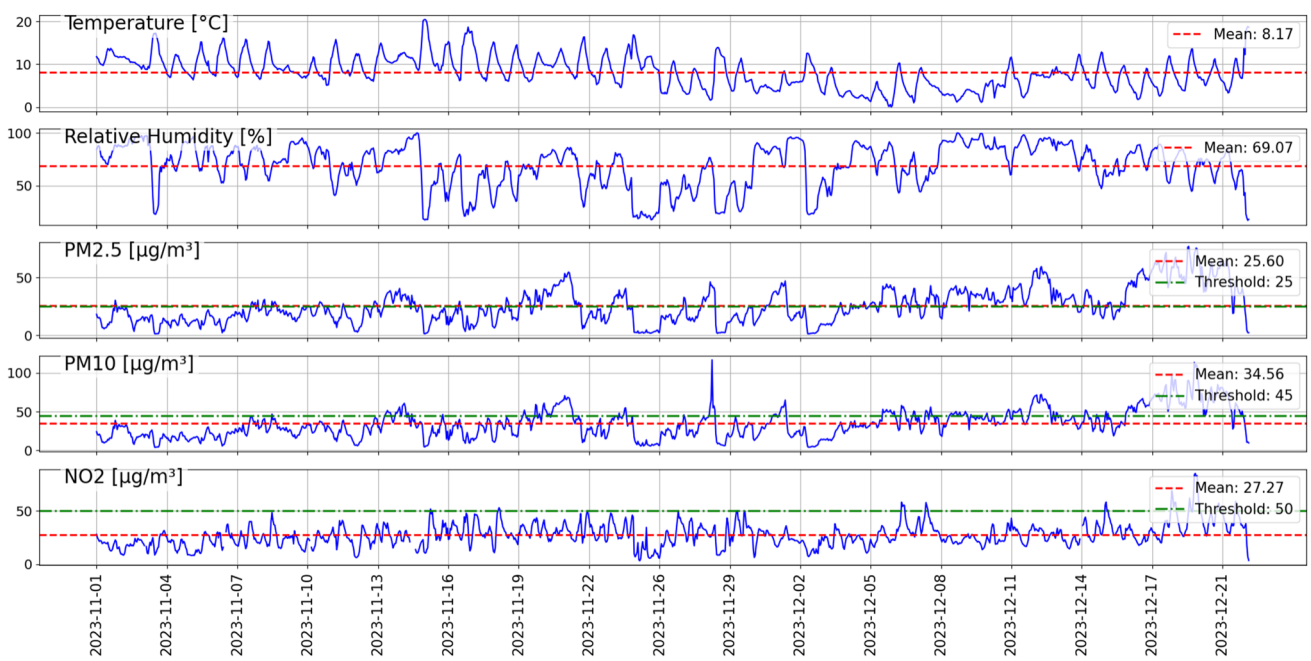


Fig. 3 Hourly trend of meteorological parameters and pollutant concentrations measured by the fixed reference station of the Politecnico di Torino. From top to bottom: Temperature, Relative Humidity, $PM_{2.5}$, PM_{10}, and NO_2. The green dashed lines in the figure represent the daily mean limit values established by the new European Directive (EU) 2024/2881

reference instruments frequently recorded concentrations above these thresholds, particularly for $PM_{2.5}$ and PM_{10} in December. While these observations highlight the critical air quality conditions of the site, a systematic analysis of the exceedances and the ability of the low-cost MONICA sensors to correctly identify them is provided in Sect. 3.4.

3.2 Performance of the Three Regression Models

The three regression models presented in Sect. 2.3 were trained using concentration data from the reference instruments and the predictor variables from the MONICA system listed in Table 1. While various training durations were tested (as discussed in the following sections), this subsection focuses on evaluating the performance of the three regression models during a standard two-week training phase, which serves as our baseline for comparison.

Figure 4 presents scatter plots comparing the hourly averaged concentrations derived from sensor measurements and those predicted by the regression models during the training phase. The RF model (third row) achieves the best overall performance, with R^2 consistently above 0.95 across all pollutants. The GAM model performs slightly worse, particularly for NO_2, with R^2 dropping to 0.66 and higher RMSE values due to underestimation of peak concentrations. MLR shows a similar behavior but with slightly lower performance compared to the other models.

During the two-week training period, the concentration ranges measured were approximately 1-40 μg/m^3 for $PM_{2.5}$, 4-60 μg/m^3 for PM_{10}, 6-53 μg/m^3 for NO_2. The regression models are thus trained on these concentration ranges.

The regression models calibrated on these two weeks, were then applied to the predictor values measured by the MONICA sensors in the following weeks. In Fig. 5, the regression model results are compared to the reference instrument measurements during Week 7 (i.e., five weeks after the training period). The scatter plots in the first row show that the MLR regression model is able to accurately reconstruct pollutant concentrations, especially for particulate matter. For both $PM_{2.5}$ and PM_{10}, data points align well with the bisector, and statistical indicators confirm a good agreement. However, NO_2 concentrations tend to be underestimated, as indicated by the negative values of the fractional bias.

In the second row, the scatter plots for the GAM model highlight a limitation in predicting high concentration values, especially for PM. At low and moderate concentrations, model performance is comparable to that of MLR, but at higher levels - beyond approximately 40 μg/m^3 for $PM_{2.5}$ and 50 μg/m^3 for PM_{10} - the predicted values increasingly deviate from the bisector. This progressive divergence reduces the R^2 and increases RMSE, indicating a deterioration in model performance. For NO_2, on the other hand, GAM predictions remain closer to the bisector across the concentration range and yield performance metrics similar to those of MLR.

The RF model (third row) exhibits a performance pattern comparable to the GAM but with a more pronounced *saturation* effect. While it accurately reproduces concentrations up to the training thresholds, it displays a characteristic flattening in the upper portion of the scatter plots. This behavior aligns with the inherent structural limitations of non-parametric models: unlike parametric approaches (e.g., MLR) that assume a global functional form, models like RF are strictly data-driven and cannot extrapolate beyond the training range e.g., (Nowack et al. 2021). Specifically, predictions are constrained to the maximum values encountered during training because the ensemble cannot predict values outside the range of its leaf nodes.

The impact of this limitation varies depending on how far test values exceed the training range. For $PM_{2.5}$ and PM_{10}, concentration levels during Week 7 reached levels nearly double the training maxima (up to 80 and 110 μg/m^3, respectively). In this regime, the superior extrapolation capability of MLR clearly highlights the saturation effect of the RF model. Conversely, for NO_2, the performance gap is less pronounced. Although the RF model remains constrained by its training ceiling, the MLR benchmark itself exhibits significant underestimation at higher concentrations. Consequently, the inherent limitations of the non-parametric model are less evident for NO_2, as the linear reference also struggles to accurately capture the extreme values in that specific range.

These results emphasize the importance of including sufficient variability, particularly higher concentration episodes, in the training dataset when calibrating low-cost sensors with non-linear data-driven approaches. At the same time, field campaigns must account for practical constraints: calibration must be conducted close to the measurement period, under similar environmental conditions, and without compromising the time allocated for actual data collection. Striking a balance between statistical robustness and operational feasibility is thus essential in designing effective calibration strategies.

While Fig. 5 focused on model performance during a specific week (Week 7), Fig. 6 illustrates how the performance of the different regression models evolves over time.

For particulate matter, the R^2 values for the GAM model drop sharply after Week 5, indicating limited robustness and a significant reduction in predictive capability over time. In contrast, MLR and RF maintain consistently high performance until around Week 8, after which a visible decrease in R^2 is observed–particularly between Weeks 8 and 9. This drop suggests that, although more stable than GAM, these

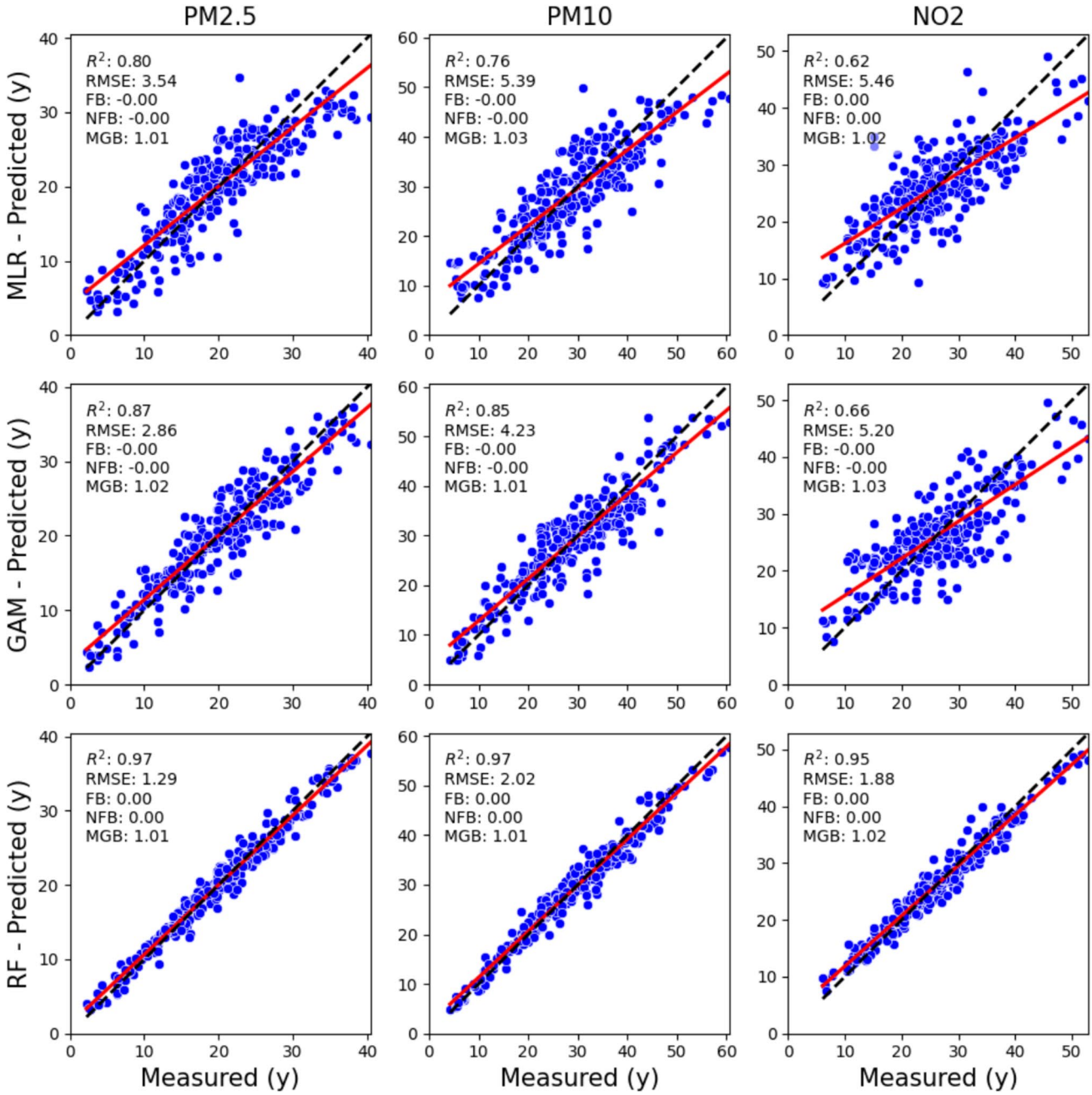


Fig. 4 Scatter plots comparing the regression results of models trained over 2 weeks with the data measured during the same 2-week period. Each column corresponds to a pollutant ($PM_{2.5}$, PM_{10}, NO_2), each row corresponds to a statistical indicator (R^2, FB, RMSE). Each point represents an hourly averaged concentration value, obtained by aggregating sensor measurements at hourly resolution

models also experience a decline in accuracy when applied beyond a certain temporal window from the calibration period. For NO_2, all three models follow a similar pattern, with R^2 values around 0.5 during the early weeks and a more pronounced decline after Week 8. This performance loss coincides with a one-month gap in reference data between Weeks 8 and 9 (see A) and likely reflects a temporal drift in sensor response. These findings suggest that, for all pollutants, model accuracy tends to degrade over time when applied to periods increasingly distant from the calibration, with NO_2 being particularly sensitive to this effect.

The RMSE trends highlight a gradual increase of error occurring immediately after the training period. For PM_{10} and $PM_{2.5}$, the error increases dramatically for the GAM model, while the MLR and the RF models maintain good performance, with MLR showing a slight advantage. Regarding NO_2, the error rises after the training period, stabilizing at a plateau before a sudden increase in Week 7.

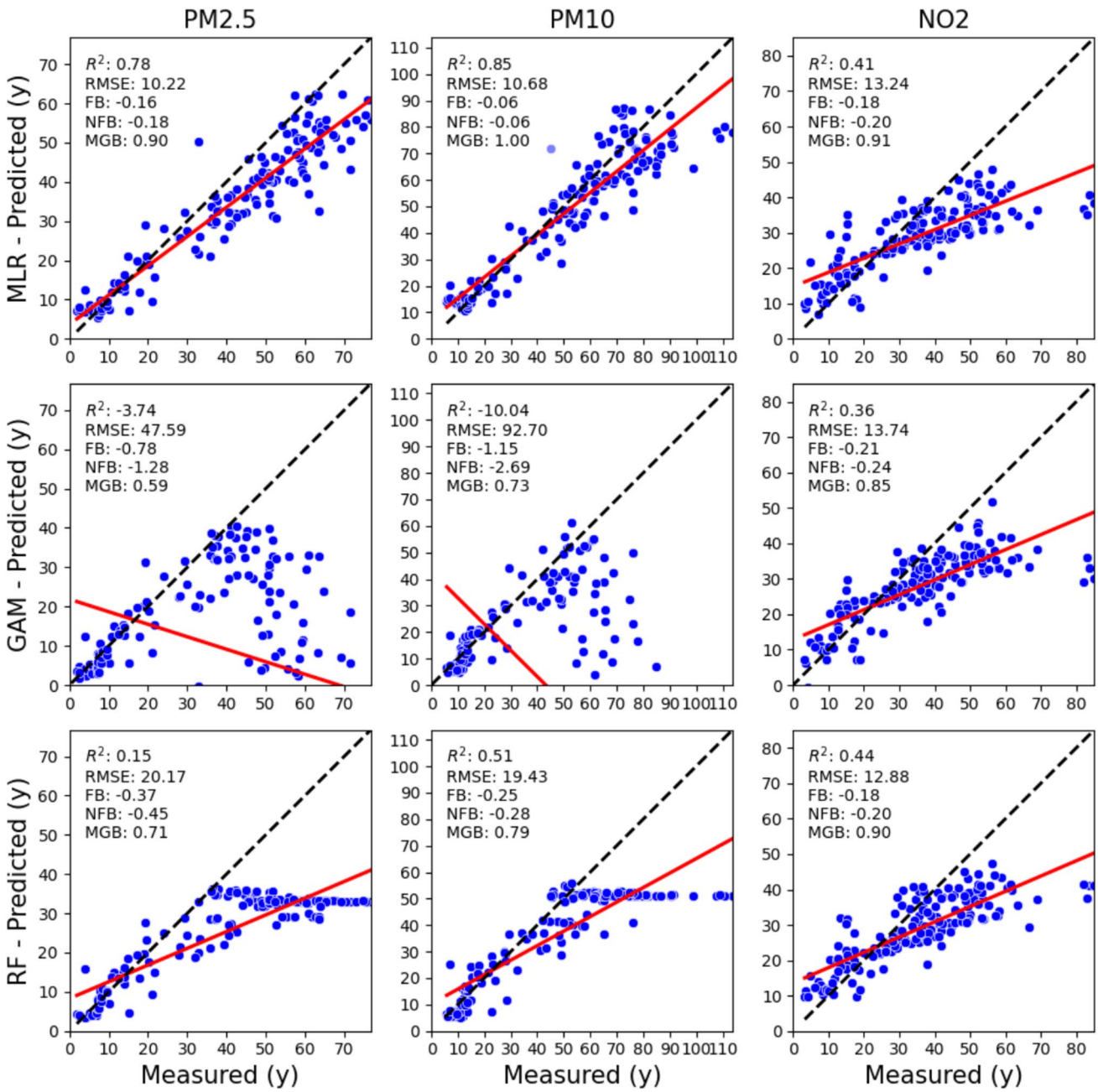


Fig. 5 Scatter plots comparing the regression results of models trained over 2 weeks with the data measured during Week 7. Each column corresponds to a pollutant ($PM_{2.5}$, PM_{10}, NO_2), each row corresponds to a statistical indicator (R^2, FB, RMSE). Each point represents an hourly median concentration value, obtained by aggregating sensor measurements at hourly resolution

Finally, the FB provides insight into whether the low-cost sensors are underestimating (negative FB) or overestimating (positive FB) pollutant concentrations compared to reference measurements. For particulate matter, the MONICA sensors progressively underestimate concentrations over time, with the GAM model performing the worst. In the case of NO_2, there is a slight overestimate in Week 4, followed by a gradual underestimate until Week 7. As highlighted also by the RMSE trends, model performance appears to improve in Week 8 (when R^2 also shows good values) but then deteriorates again in Week 9 (after the one-month interruption).

Overall, Fig. 6 confirms that the MLR model consistently provides the most reliable results across pollutants and time, though even this model shows performance degradation when applied far from the calibration window.

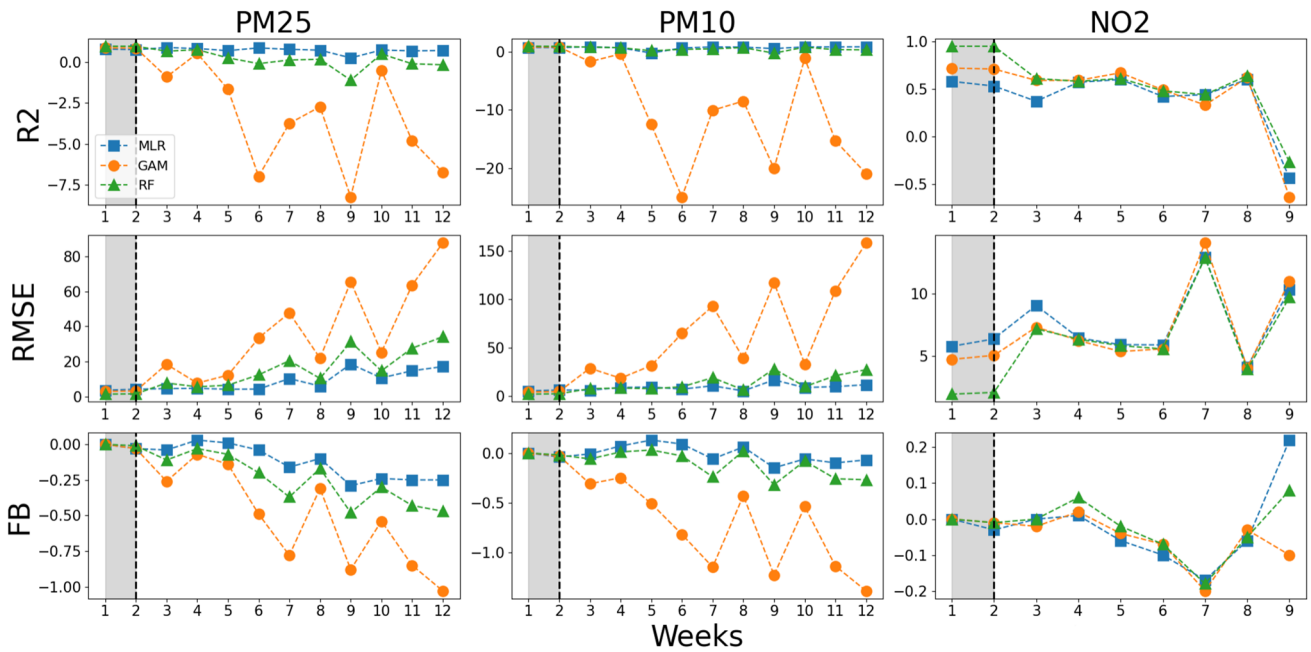


Fig. 6 Evolution of performance metrics for the different models (MLR, GAM, and RF) considering 2 weeks of training (shaded area): calibration performance over the weeks. Each column corresponds to a pollutant ($PM_{2.5}$, PM_{10}, NO_2). The three regression models–MLR, GAM, and RF–are represented by blue squares, orange circles, and green triangles, respectively. The rows display the trends for R^2, RMSE, and FB over time, illustrating how model performance evolves

The same analysis using a 3-week calibration period is provided in Appendix B, while the choice of training duration is further discussed in the following section.

3.3 Choice of the Calibration Duration

In the previous section, the performance of the regression models was analyzed using a two-week training period. This duration was selected based on a specific analysis examining how the length of the calibration phase affects the predictive accuracy of the low-cost sensors.

Figure 7 presents this analysis for the MLR model, identified as the most reliable model overall. Different markers represent the temporal evolution of model performance for varying training durations, ranging from 1 to 4 weeks. Notably, the curve corresponding to a two-week training period (orange circles) aligns with the MLR results shown in Fig. 6. The training weeks were selected as the initial N weeks of the campaign to simulate a realistic deployment scenario, where a sensor is typically calibrated immediately after installation.

For all three performance metrics (R^2, RMSE, and FB), increasing the training period from one to two weeks significantly enhances the accuracy of low-cost sensor measurements for $PM_{2.5}$ and PM_{10}. However, for NO_2, the analysis suggests that even a one-week training period is sufficient, as extending the training duration beyond that does not lead to substantial performance improvements.

This difference in optimal training duration may be explained by the representativeness of the concentration ranges encountered during the initial weeks. For NO_2, the first week alone was characterized by a high dynamic range (6–48 μg/m^3), which already covered 57% of the maximum concentration observed over the entire three-month campaign (85 μg/m^3). In contrast, for $PM_{2.5}$ and PM_{10}, the first week captured a more limited portion of the seasonal peaks (23% and 35%, respectively). By extending the training to two weeks, these coverage percentages increased to 31% (40 μg/m^3) for $PM_{2.5}$ and 46% (60 μg/m^3) for PM_{10}.

Compared to Figs. 6 and 7 provides a clearer depiction of the long-term calibration stability for the MLR model, refining the findings from the previous section.

For particulate matter, both RMSE and FB show a progressive decline in performance starting from Week 6, with a more marked degradation from Week 9 onward–following the one-month interruption in the measurements. However, R^2 remains relatively high even after this point (see values from Week 10 to 12). This indicates that while the model continues to capture the temporal variability and overall trends in PM concentrations (i.e., good correlation), it progressively diverges in terms of absolute values–resulting in increased bias and error. In other words, the model maintains its ability to follow the shape of the time series, but its accuracy in estimating the actual concentration levels decreases over time.

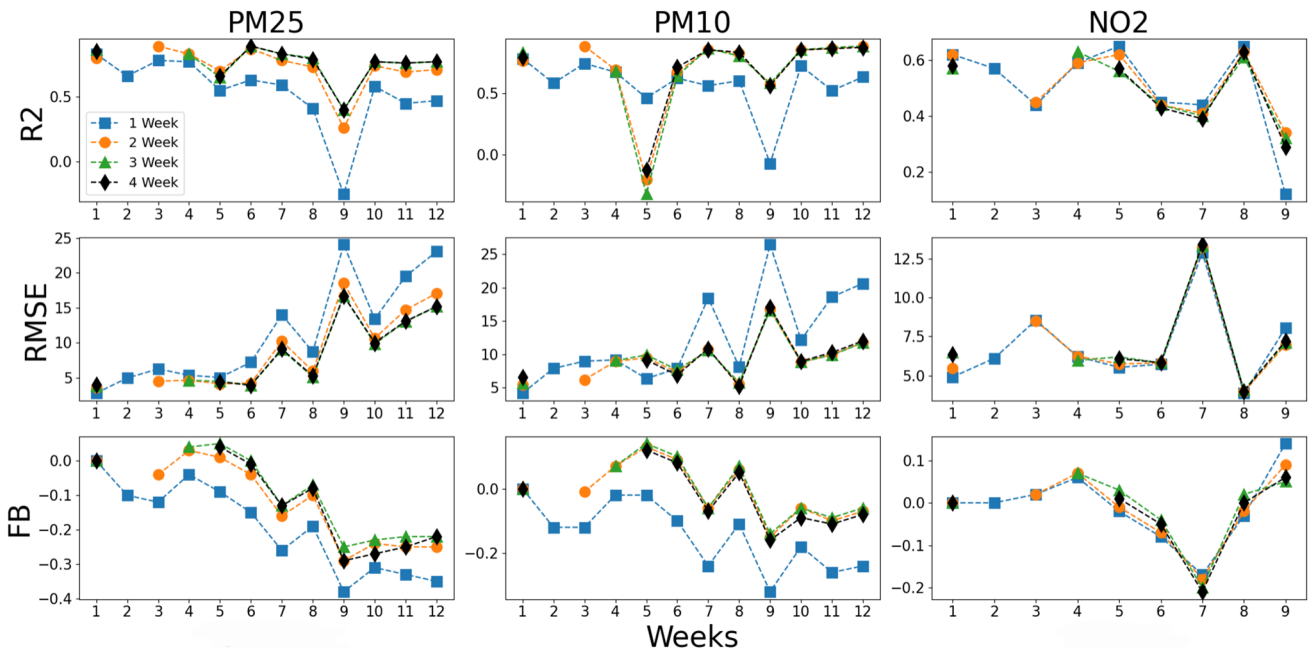


Fig. 7 Evolution of performance metrics for MLR model considering different weeks as training: effect of calibration duration on performance. Each column corresponds to a pollutant ($PM_{2.5}$, PM_{10}, NO_2), each row corresponds to a statistical indicator (R2, FB, RMSE). Each panel represents the trend of the statistical indicator over the weeks for each model used

In contrast, the performance for NO_2 does not exhibit a consistent trend but shows a localized increase in error and bias in Week 7, and a sharp drop in all the three statistics in Week 9–after the one-month interruption. This may indicate a shift in the underlying pollutant dynamics that is not accounted for by the initial calibration, highlighting the need for more frequent recalibration in the case of NO_2, a reactive gas with greater temporal variability.

Overall, this analysis suggests that the long-term stability of calibration models can vary significantly across pollutants and performance metrics. While the MLR model maintains good correlation with reference data over time–particularly for particulate matter–its accuracy in predicting absolute concentrations tends to deteriorate. For NO_2, calibration stability is more sensitive to time and environmental variability, indicating the need for more frequent recalibration to ensure reliable performance.

This sensitivity is closely linked to the seasonal variability of pollutants and meteorological drivers. The performance degradation observed after the one-month interruption suggests that short-term training may not fully capture the shifts in chemical and physical dynamics that emerge as seasonal patterns evolve.

3.4 Daily Concentration Trends and Regulatory Threshold Analysis

After identifying the most suitable regression model (i.e., MLR), determining the optimal calibration duration (i.e., two weeks), and evaluating its long-term performance, we examined the temporal evolution of daily-averaged concentrations for $PM_{2.5}$ (Fig. 8), PM_{10} (Fig. 9), and NO_2 (Fig. 10). This analysis is particularly relevant in light of the new Directive (EU) 2024/2881, which introduces stricter daily (24-hour) limit values compared to previous legislation. Specifically, a new daily threshold of 50 μg/m^3 is established for NO_2 (not to be exceeded more than 18 times per calendar year), whereas the previous directive focused primarily on hourly peaks. Regarding particulate matter, the daily limit for PM_{10} has been reduced from 50 μg/m^3 to 45 μg/m^3, and the allowed number of annual exceedances has been halved from 35 to 18. Additionally, for $PM_{2.5}$, a daily limit of 25 μg/m^3 has been introduced (not to be exceeded more than 18 times per year)–a parameter that was not strictly regulated in the 2008/50/EC Directive. These changes reflect a more rigorous approach to urban air quality management.

Figures 8, 9, and 10 illustrate the daily-averaged concentrations for $PM_{2.5}$, PM_{10}, and NO_2 throughout the campaign, with the interruption period of the measurements indicated by two wavy vertical lines. For each pollutant, the plots indicate the relevant daily concentration limits set by the new EU Directive. In addition to pollutant concentrations, the figures also display the trend of relative humidity, the only meteorological variable found to significantly contribute to the site-specific calibration models (see Sect. 2.3).

Figure 8 illustrates the progressive increase in $PM_{2.5}$ concentrations throughout the winter period. Measurements

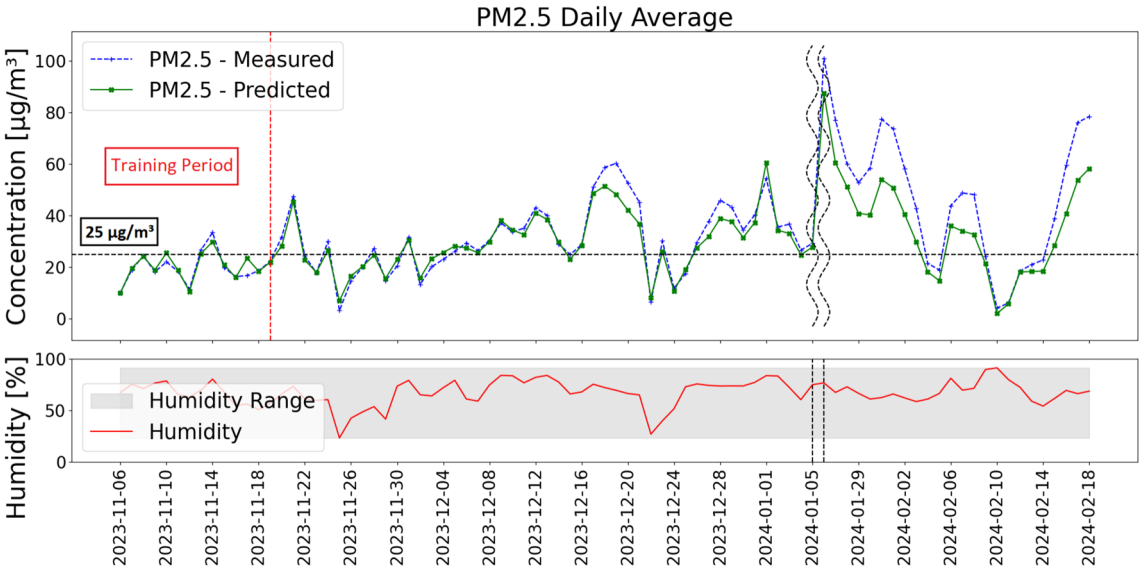


Fig. 8 Comparison of daily average $PM_{2.5}$ concentrations measured by the reference instrument (dashed blue line) and those estimated by MONICA (green line) using the MLR regression model trained over a two-week period. The regulatory limit established by Directive (EU) 2024/2881 is shown as a dashed horizontal line. The trend of relative humidity is also reported (red line). The measurement interruption period is indicated by vertical dashed lines

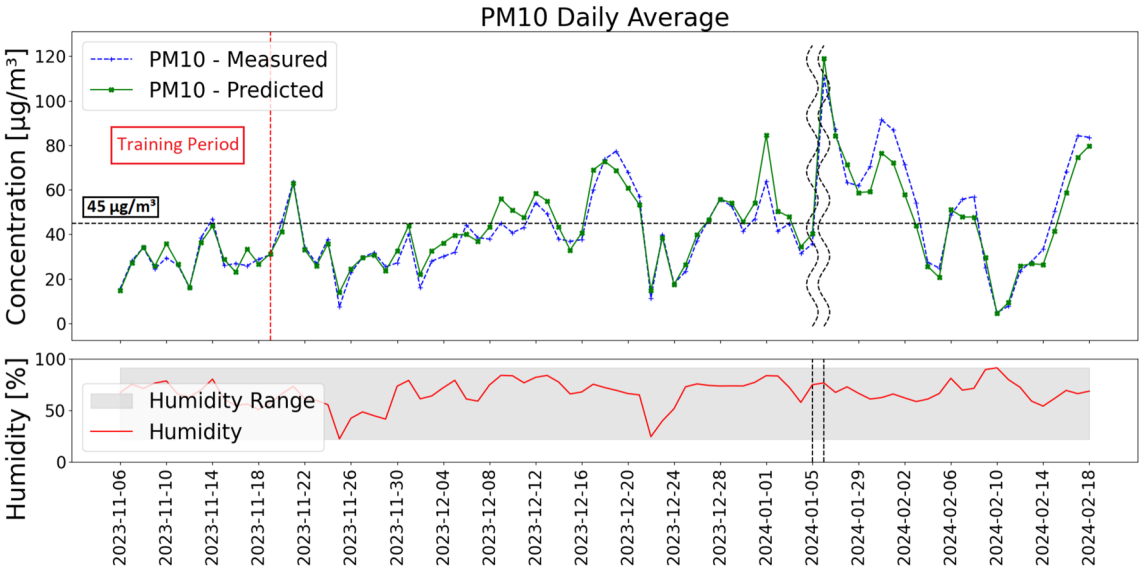


Fig. 9 Comparison of daily average PM_{10} concentrations measured by the reference instrument (dashed blue line) and those estimated by MONICA (green line) using the MLR regression model trained over a two-week period. The regulatory limit established by Directive (EU) 2024/2881 is shown as a dashed horizontal line. The trend of relative humidity is also reported (red line). The measurement interruption period is indicated by vertical dashed lines

from the MONICA sensors show good agreement with those from the reference instruments. As previously discussed, a gradual divergence between the two datasets is observed, likely due to the progressive degradation of calibration performance. Multiple exceedances of the 25 µg/m^3 daily concentration limit, as established by Directive (EU) 2024/2881, are recorded.

The agreement between MONICA and reference instruments is even more consistent for PM_{10} (Fig. 9), reflecting the slightly lower RMSE and FB values obtained for PM_{10} compared to $PM_{2.5}$ (see Fig. 7). The temporal trends of PM_{10} and $PM_{2.5}$ are closely aligned. Notably, exceedances of the 45 µg/m^3 daily limit are frequent starting from the second week of December. In Fig. 10, MONICA sensors demonstrate a strong ability to replicate both the temporal dynamics and concentration levels of NO_2. However, a systematic underestimation of extreme values–both peaks and troughs–is evident. Notably, a significant underestimation

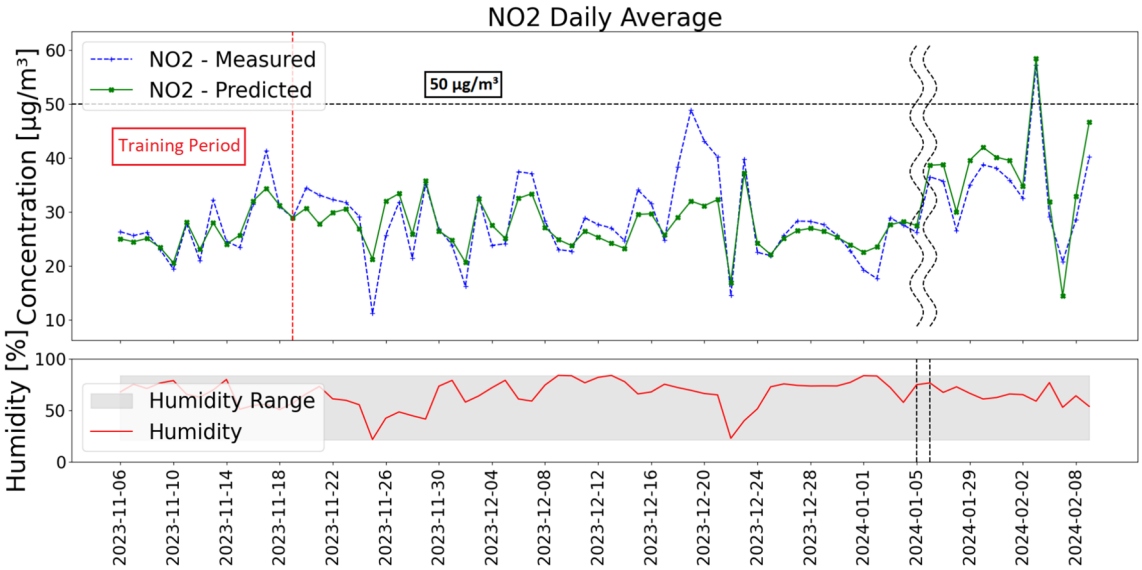


Fig. 10 Comparison of daily average NO_2 concentrations measured by the reference instrument (dashed blue line) and those estimated by MONICA (green line) using the MLR regression model trained over a two-week period. The regulatory limit established by Directive (EU) 2024/2881 is shown as a dashed horizontal line. The trend of relative humidity is also reported (red line). The measurement interruption period is indicated by vertical dashed lines

Table 4 Number of exceedances based on the daily limits established by European Directive (EU) 2024/2881, as estimated over a 12-week measurement period for particulate matter (PM_{10} and $PM_{2.5}$) and a 9-week measurement period for NO_2, along with average concentrations over the entire period for each pollutant

Instrument	Exceedances [#]			Average concentration		
	$PM_{2.5}$ 25 µg/m^3	PM_{10} 45 µg/m^3	NO_2 50 µg/m^3	$PM_{2.5}$ [µg/m^3]	PM_{10} [µg/m^3]	NO_2 [µg/m^3]
Ref. Instr	51	31	1	33.90	42.43	29.26
MONICA	50	33	1	27.90	36.81	29.02

occurs between 16 and 20 December (Week 7). This period is marked by a peak in RMSE and a negative FB, as shown in Fig. 6. After the interruption period, while R^2 in Fig. 6 shows a significant decline for Week 9, a clear decrease in correlation is not evident in the time series data. Similar to particulate matter, NO_2 levels increase over the study period, peaking in early February, a trend that is well captured by the low-cost sensor. However, unlike PM, NO_2 concentrations exceed the regulatory threshold of 50 µg/m^3 only once throughout the entire period.

As a complementary analysis, Appendix B reports the same temporal trends of Figs. 8, 9, and 10 using a three-week calibration period. However, no substantial differences were observed compared to the two-week calibration, confirming its adequacy for this application.

Table 4 summarizes the number of exceedances for each pollutant, based on the daily thresholds set by Directive (EU) 2024/2881. The number of exceedances estimated by MONICA (two-weeks training) is almost identical to that recorded by reference instrumentation, underscoring the potential of low-cost sensors for regulatory monitoring purposes. Notably, the daily limit for $PM_{2.5}$ is exceeded on more than 60% of the days in the observation period, approximately 40% for PM_{10}, and around 2% for NO_2. Directive (EU) 2024/2881 allows a maximum of 18 exceedances per year for each pollutant; this threshold is significantly surpassed in our dataset, despite the dataset spanning slightly more than three months.

Table 4 also includes the average concentrations for each pollutant over the measurement period. As expected, given the predominantly negative fractional bias values reported earlier, the MONICA sensors tend to underestimate concentrations by approximately 18% for $PM_{2.5}$, 13% for PM_{10}, and 1% for NO_2. It is worth noting that the average $PM_{2.5}$ concentration measured during the campaign exceeds the current annual limit of 10 µg/m^3 set by Directive (EU) 2024/2881. A similar pattern is observed for PM_{10}, whose reference average of 42.43 µg/m^3 exceeds both the current (European Directive 2008/50/CE (Parliament and Council 2008)) annual limit of 40 µg/m^3 and the upcoming 20 µg/m^3 limit. For NO_2, only the newly introduced annual limit of 20 µg/m^3 is exceeded. In performing these comparison, it is worth remembering that the measurements reported here refer only to the winter season, during which meteorological

conditions (and increased pollutant emissions, notably due to domestic heating) typically lead to the highest pollutant levels observed over the whole year.

4 Conclusions

This study evaluated the performance and calibration strategies of the MONICA low-cost sensor system for measuring particulate matter ($PM_{2.5}$ and PM_{10}) and nitrogen dioxide (NO_2) in an urban setting. Our primary objective was to identify the most reliable calibration model, assess the optimal co-location duration, and analyze the long-term stability of sensor performance.

Among the models tested, Multiple Linear Regression (MLR) proved to be significantly more robust than both the Generalized Additive Model (GAM) and Random Forest (RF) models. The latter models exhibited high sensitivity to the dataset used for calibration and struggled to predict the highest concentration values accurately.

In terms of calibration duration, our analysis revealed that a two-week calibration period was sufficient for particulate matter, providing reliable sensor performance for up to seven weeks post-calibration. For NO_2, however, a shorter, one-week calibration period proved to be adequate, as extending the duration beyond that did not result in significant performance improvements. This suggests that the calibration duration can vary depending on the pollutant and, more importantly, on the environmental representativeness of the training window. In our study, the high urban variability allowed the sensors to encounter a significant portion of the seasonal dynamic range (covering up to 57% of the campaign maximum for NO_2 in the first week, and 46% for PM_{10} by the second) within these short timeframes. This ensured that the models were well-constrained from the outset for the specific conditions of the monitored period.

The long-term evaluation revealed important differences in sensor behavior. While PM sensors maintained a good correlation with reference data over time, their accuracy in predicting absolute concentrations declined gradually, as reflected in increasing RMSE and fractional bias. For NO_2, a more abrupt degradation was observed, especially after two months, suggesting a higher susceptibility to sensor drift and environmental variability. This degradation also reflects the inherent seasonal limitations of the study; while the models included parameters to compensate for ozone interference, the narrow range of conditions typical of a winter campaign may limit their applicability to summer months, when higher ozone levels and broader temperature fluctuations occur. A general tendency toward underestimation was found for all pollutants, reinforcing the need for regular validation against reference-grade instruments.

Despite these limitations, MONICA sensors successfully captured temporal trends and regulatory exceedances, confirming their suitability for urban air quality surveillance and public health applications. Their affordability, compact design, and acceptable long-term performance make them attractive tools for enhancing spatial and temporal resolution in monitoring networks.

In summary, this study demonstrates that low-cost multi-pollutant systems like MONICA can deliver reliable results over extended periods, provided that calibration strategies are carefully designed and periodically updated. The findings provide operational guidance for optimizing sensor deployments under real-world constraints, contributing to the development of scalable and cost-effective air quality networks.

Future work should explore multi-seasonal deployments, longer observation periods, and parallel sensor configurations to evaluate spatial variability and further refine calibration methods. Integrating these insights into policy and urban planning could enhance the responsiveness and equity of air quality management strategies.

Time Windows with Valid Data for Calibration and Evaluation

The monitoring campaign took place from November 6, 2023, to February 25, 2024, with MONICA sensors co-located alongside reference instruments. Due to interruptions in the reference instruments, the time windows with valid data for direct comparison were limited: 12 full weeks for PM2.5 and PM10, and 9 full weeks for NO_2. Table A.1

Table 5 Availability of valid data for $PM_{2.5}$, PM_{10}, and NO_2, broken down by full weeks (Monday to Sunday)

Week	Date range	Available pollutants
Week 1	06–12 Nov 2023	$PM_{2.5}$, PM_{10}, NO_2
Week 2	13–19 Nov 2023	$PM_{2.5}$, PM_{10}, NO_2
Week 3	20–26 Nov 2023	$PM_{2.5}$, PM_{10}, NO_2
Week 4	27 Nov – 03 Dec 2023	$PM_{2.5}$, PM_{10}, NO_2
Week 5	04–10 Dec 2023	$PM_{2.5}$, PM_{10}, NO_2
Week 6	11–17 Dec 2023	$PM_{2.5}$, PM_{10}, NO_2
Week 7	18–24 Dec 2023	$PM_{2.5}$, PM_{10}, NO_2
Week 8	25–31 Dec 2023	$PM_{2.5}$, PM_{10}, NO_2
–	01–07 Jan 2024	$PM_{2.5}$, PM_{10}, NO_2 (only until 05 Jan)
–	08–14 Jan 2024	No data
–	15–21 Jan 2024	No data
–	22–28 Jan 2024	$PM_{2.5}$, PM_{10}, NO_2 (starting 25 Jan)
Week 9	29 Jan – 04 Feb 2024	$PM_{2.5}$, PM_{10}, NO_2
Week 10	05–11 Feb 2024	$PM_{2.5}$, PM_{10}
Week 11	12–18 Feb 2024	$PM_{2.5}$, PM_{10}
Week 12	19–25 Feb 2024	$PM_{2.5}$, PM_{10}

Only weeks with complete data coverage are numbered

indicates the weeks with complete data coverage as well as those with partial or no data.

Results for 3-Weeks Training

The comparison of three regression models and the analysis of their performance relative to the training period indicated that a 2-week training period for the MLR model offers the optimal balance for real-world campaigns conducted within a single season. In this appendix, we present the comparison between MONICA measurements and those from the reference instruments using a 3-week training period for the MLR model.

Figures 11, 12, and 13 show that a 3-week training period yields results that are practically identical to those observed in Figs. 9 and 10, confirming that a 2-week period is sufficient and optimal for calibration.

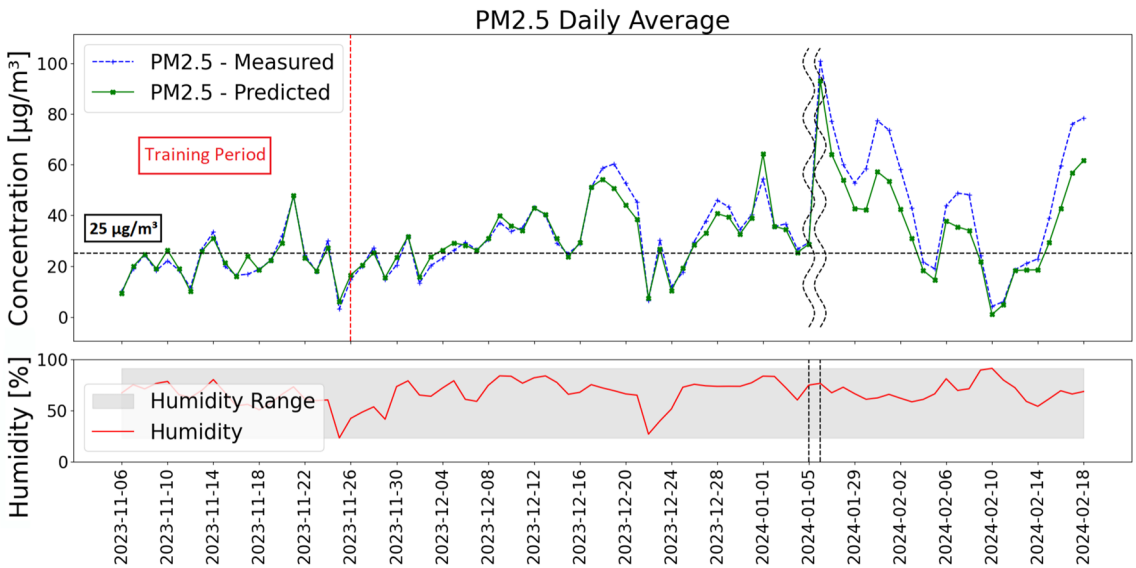


Fig. 11 Comparison of daily average $PM_{2.5}$ concentrations measured by the reference instrument (dashed blue line) and those estimated by MONICA (green line) using the MLR regression model trained over a three-week period. The regulatory limit established by Directive (EU) 2024/2881 is shown as a dashed horizontal line. The trend of relative humidity is also reported (red line). The measurement interruption period is indicated by vertical dashed lines

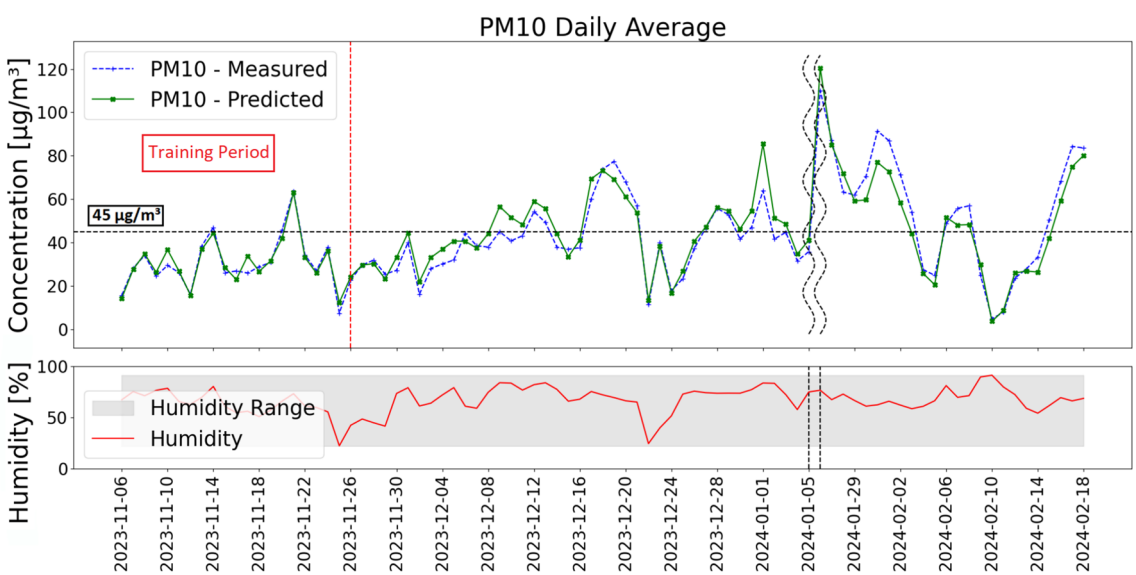


Fig. 12 Comparison of daily average PM_{10} concentrations measured by the reference instrument (dashed blue line) and those estimated by MONICA (green line) using the MLR regression model trained over a three-week period. The regulatory limit established by Directive (EU) 2024/2881 is shown as a dashed horizontal line. The trend of relative humidity is also reported (red line). The measurement interruption period is indicated by vertical dashed lines

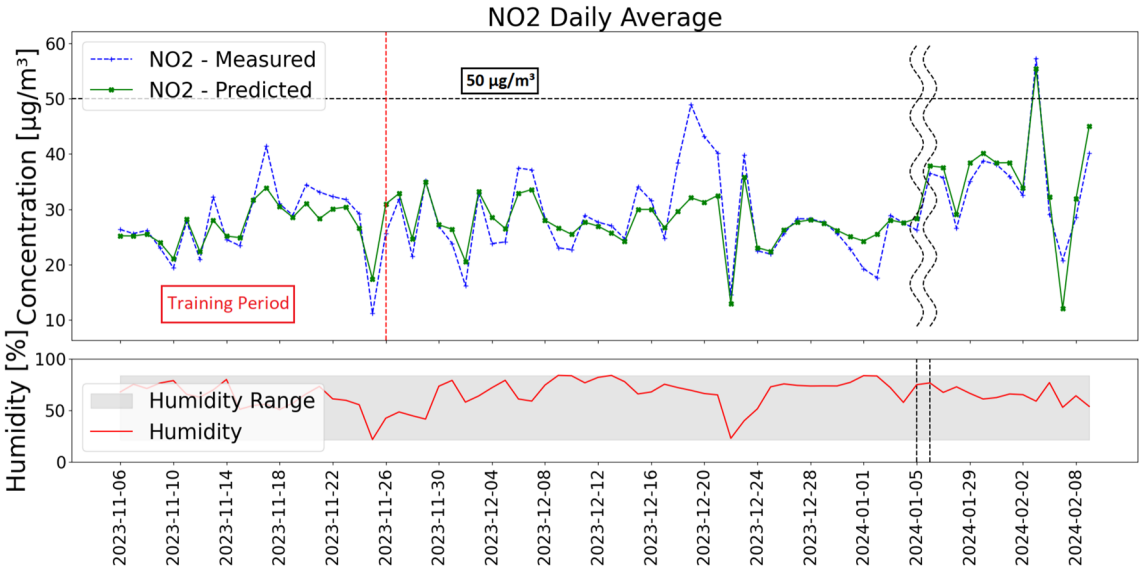


Fig. 13 Comparison of daily average NO_2 concentrations measured by the reference instrument (dashed blue line) and those estimated by MONICA (green line) using the MLR regression model trained over a three-week period. The regulatory limit established by Directive (EU) 2024/2881 is shown as a dashed horizontal line. The trend of relative humidity is also reported (red line). The measurement interruption period is indicated by vertical dashed lines

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s44408-026-00150-1.

Acknowledgements This work was conducted within the climate_transition@polito project, funded by the Italian Ministry of University and Research under the 2023–2027 "Departments of Excellence" program. The authors thank Gerardo D'Elia, Saverio De Vito, and Sergio Ferlito (ENEA – Portici Research Center) for their technical support and collaboration. P.S. acknowledges funding from the Auvergne-Rhône-Alpes Region through the VEB_{MU} project. S.F. and M.C. acknowledge funding from Fondazione Compagnia di San Paolo through the LIBRA project.

Author contributions S.F. conceptualized the study, developed the methodology, and wrote the original draft. D.G. curated data, created visualizations, and contributed to writing the original draft. V.V. performed formal analysis, created visualizations, and contributed to writing the original draft. N.M. curated data and helped review the manuscript. G.F. provided resources. M.C. conceptualized the study, supervised and managed the project, reviewed the manuscript. P.S. conceptualized the study, supervised the work, reviewed the manuscript, and managed the project.

Funding The authors did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors for the research, authorship, and/or publication of this article.

Data availability The dataset generated and analyzed during the current study, including raw low-cost sensor outputs and reference instrument data, is provided as Supplementary Material to ensure the transparency and reproducibility of the results.

Declarations

Competing interests The authors declare no conflict of interest.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, Dominici F (2017) Association of short-term exposure to air pollution with mortality in older adults. JAMA 318(24):2446–2456

Hystad P, Larkin A, Rangarajan S, AlHabib KF, Avezum Á, Calik KBT, Chifamba J, Dans A, Diaz R, Du Plessis JL et al (2020) Associations of outdoor fine particulate air pollution and cardiovascular disease in 157 436 individuals from 21 high-income, middle-income, and low-income countries (pure): a prospective cohort study. Lancet Planetary Health 4(6):235–245

Lin C-K, Chang Y-T, Lee F-S, Chen S-T, Christiani D (2021) Association between exposure to ambient particulate matters and risks of autism spectrum disorder in children: a systematic review and exposure-response meta-analysis. Environ Res Lett 16(6):063003

De Bont J, Krishna B, Stafoggia M, Banerjee T, Dholakia H, Garg A, Ingole V, Jaganathan S, Kloog I, Lane K et al (2024) Ambient air pollution and daily mortality in ten cities of india: a causal modelling study. The Lancet Planetary Health 8(7):433–440

Salizzoni P, Soulhac L, Mejean P (2009) Street canyon ventilation and atmospheric turbulence. Atmos Environ 43(32):5056–5067

Fellini S, Salizzoni P, Ridolfi L (2021) Vulnerability of cities to toxic airborne releases is written in their topology. Sci Rep 11(1):23029

Kim K-H, Kabir E, Kabir S (2015) A review on the human health impact of airborne particulate matter. Environ Int 74:136–143. https://doi.org/10.1016/j.envint.2014.10.005

Chen J, Hoek G (2020) Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. Environ Int 143:105974. https://doi.org/10.1016/j.envint.2020.105974

Kang J, Choi K (2024) Calibration methods for low-cost particulate matter sensors considering seasonal variability. Sensors 24(10):3023. https://doi.org/10.3390/s24103023

Pope CA, Dockery DW (2006) Health effects of fine particulate air pollution: lines that connect. J Air Waste Manag Associat 56(6):709–742

WHO: Review of evidence on health aspects of air pollution: Revihaap project: technical report. In: Review of Evidence on Health Aspects of Air Pollution: REVIHAAP Project: Technical Report, (2021)

Khreis H, Kelly C, Tate J, Parslow R, Lucas K, Nieuwenhuijsen M (2017) Exposure to traffic-related air pollution and risk of development of childhood asthma: a systematic review and meta-analysis. Environ Int 100:1–31

Anenberg SC, Mohegh A, Goldberg DL, Kerr GH, Brauer M, Burkart K, Hystad P, Larkin A, Wozniak S, Lamsal L (2022) Long-term trends in urban no_2 concentrations and associated paediatric asthma incidence: estimates from global datasets. Lancet Planetary Health 6(1):49–58

Soulhac L, Fellini S, Nguyen CV, Salizzoni P (2023) Evaluation of photostationary and non-photostationary operational models for nox pollution in a street canyon. Atmos Environ 297:119589

European Parliament and Council: Directive (EU) 2024/2881 of the European Parliament and of the Council of 23 October 2024 on ambient air quality and cleaner air for Europe. https://eur-lex.europa.eu/eli/dir/2024/2881/oj/eng (2024)

Steinle S, Reis S, Sabel CE, Semple S, Twigg MM, Braban CF, Leeson SR, Heal MR, Harrison D, Lin C, Wu H (2015) Personal exposure monitoring of PM 2.5 in indoor and outdoor microenvironments. Sci Total Environ 508:383–394. https://doi.org/10.1016/j.scitotenv.2014.12.003

Fellini S, Salizzoni P, Ridolfi L (2024) Effect of emission variability on concentration fluctuations in idealised deep urban street canyons. Urban Clim 55:101952

McHugh C, Carruthers D, Edmunds H (1997) ADMS-Urban: an air quality management system for traffic, domestic and industrial pollution. Int J Environ Pollut 8(3–6):666–674

Soulhac L, Salizzoni P, Cierco F, Perkins R (2011) The model SIRANE for atmospheric urban pollutant dispersion; part i, presentation of the model. Atmos Environ 45(39):7379–7395

Kim Y, Wu Y, Seigneur C, Roustan Y (2018) Multi-scale modeling of urban air pollution: development and application of a street-in-grid model (v1. 0) by coupling munich (v1. 0) and polair3d (v1. 8.1). Geosci Model Dev 11(2):611–629

Suter I, Grylls T, Sützl B, Reeuwijk M (2021) a large-eddy-simulation model for urban environments. Geosci Model Dev Discuss 2021:1–40

Hoek G, Beelen R, De Hoogh K, Vienneau D, Gulliver J, Fischer P, Briggs D (2008) A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 42(33):7561–7578

Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, Tsai M-Y, Künzli N, Schikowski T, Marcon A et al (2013) Development of no2 and nox land use regression models for estimating air pollution exposure in 36 study areas in europe-the escape project. Atmos Environ 72:10–23

Ma X, Zou B, Deng J, Gao J, Longley I, Xiao S, Guo B, Wu Y, Xu T, Xu X et al (2024) A comprehensive review of the development of land use regression approaches for modeling spatiotemporal variations of ambient air pollution: A perspective from 2011 to 2023. Environ Int 183:108430

Liu S, Tian H, Luo L, Bai X, Zhu C, Lin S, Zhao S, Zhang K, Hao J, Guo Z, Lv Y (2022) Health impacts and spatiotemporal variations of fine particulate and its typical toxic constituents in five urban agglomerations of China. Sci Total Environ 806:151459. https://doi.org/10.1016/j.scitotenv.2021.151459

Schmitz S, Caseiro A, Kerschbaumer A, Von Schneidemesser E (2024) Quantifying impacts of local traffic policies on PM concentrations using low cost sensors in berlin. Aerosol Air Quality Res 24(9):240050. https://doi.org/10.4209/aaqr.240050

Tang H, Cai Y, Gao S, Sun J, Ning Z, Yu Z, Pan J, Zhao Z (2024) Multi-scenario validation and assessment of a particulate matter sensor monitor optimized by machine learning methods. Sensors 24(11):3448. https://doi.org/10.3390/s24113448

Cross ES, Williams LR, Lewis DK, Magoon GR, Onasch TB, Kaminsky ML, Worsnop DR, Jayne JT (2017) Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. Atmospheric Measur Techn 10(9):3575–3588. https://doi.org/10.5194/amt-10-3575-2017

Thomas GW, Sousan S, Tatum M, Liu X, Zuidema C, Fitzpatrick M, Koehler KA, Peters TM (2018) Low-cost, distributed environmental monitors for factory worker health. Sensors 18(5):1411. https://doi.org/10.3390/s18051411. (**Accessed 2024-12-09**)

Johnson KK, Bergin MH, Russell AG, Hagler GSW (2018) Field test of several low-cost particulate matter sensors in high and low concentration urban environments. Aerosol Air Quality Res 18(3):565–578. https://doi.org/10.4209/aaqr.2017.10.0418. (**Accessed 2024-12-09**)

Hagan DH, Isaacman-VanWertz G, Franklin JP, Wallace LMM, Kocar BD, Heald CL, Kroll JH (2018) Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. Atmospheric Measur Techn 11(1):315–328. https://doi.org/10.5194/amt-11-315-2018

Badura M, Batog P, Drzeniecka-Osiadacz A, Modzel P (2018) Evaluation of low-cost sensors for ambient pm2.5 monitoring. J Sens 1:5096540. https://doi.org/10.1155/2018/5096540

Cavaliere A, Carotenuto F, Di Gennaro F, Gioli B, Gualtieri G, Martelli F, Matese A, Toscano P, Vagnoli C, Zaldei A (2018) Development of low-cost air quality stations for next generation monitoring networks: Calibration and validation of pm2.5 and pm10 sensors. Sensors 18(9):2843. https://doi.org/10.3390/s18092843

Wang Y, Li J, Jing H, Zhang Q, Jiang J, Biswas P (2015) Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement. Aerosol Sci Technol 49(11):1063–1077

Bulot FMJ, Johnston SJ, Basford PJ, Easton NHC, Apetroaie-Cristea M, Foster GL, Morris AKR, Cox SJ, Loxham M (2019) Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. Sci Rep 9(1):7497. https://doi.org/10.1038/s41598-019-43716-3. (**Accessed 2024-12-09**)

Giordano MR, Malings C, Pandis SN, Presto AA, McNeill VF, Westervelt DM, Beekmann M, Subramanian R (2021) From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors. J Aerosol Sci 158:105833. https://doi.org/10.1016/j.jaerosci.2021.105833

Venkatraman Jagatha J, Klausnitzer A, Chacón-Mateos M, Laquai B, Nieuwkoop E, Van Der Mark P, Vogt U, Schneider C (2021) Calibration method for particulate matter low-cost sensors used in ambient air quality monitoring and research. Sensors 21(12):3960. https://doi.org/10.3390/s21123960. (**Accessed 2025-01-27**)

Zimmerman N (2022) Tutorial: Guidelines for implementing low-cost sensor networks for aerosol monitoring. J Aerosol Sci 159:105872. https://doi.org/10.1016/j.jaerosci.2021.105872

Crilley LR, Shaw M, Pound R, Kramer LJ, Price R, Young S, Lewis AC, Pope FD (2018) Evaluation of a low-cost optical particle counter (alphasense opc-n2) for ambient air monitoring. Atmospheric Measur Tech 11(2):709–720. https://doi.org/10.5194/amt-11-709-2018

Di Antonio A, Popoola OAM, Ouyang B, Saffell J, Jones RL (2018) Developing a relative humidity correction for low-cost sensors measuring ambient particulate matter. Sensors 18(9):2790. https://doi.org/10.3390/s18092790. (**Accessed 2025-01-30**)

Macías-Hernández BA, Tello-Leal E, Barrios SO, Leiva-Guzmán MA, Toro AR (2023) Effect of environmental conditions on the performance of a low-cost atmospheric particulate matter sensor. Urban Clim 52:101753. https://doi.org/10.1016/j.uclim.2023.101753

Datta A, Saha A, Zamora ML, Buehler C, Hao L, Xiong F, Gentner DR, Koehler K (2020) Statistical field calibration of a low-cost PM2.5 monitoring network in Baltimore. Atmos Environ 242:117761. https://doi.org/10.1016/j.atmosenv.2020.117761

Jiao W, Hagler G, Williams R, Sharpe R, Brown R, Garver D, Judge R, Caudill M, Rickard J, Davis M, Weinstock L, Zimmer-Dauphinee S, Buckley K (2016) Community air sensor network (cairsense) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern united states. Atmospheric Measur Tech 9(11):5281–5292. https://doi.org/10.5194/amt-9-5281-2016

Mijling B, Jiang Q, De Jonge D, Bocconi S (2018) Field calibration of electrochemical NO_2 sensors in a citizen science context. Atmospheric Measur Tech 11(3):1297–1312. https://doi.org/10.5194/amt-11-1297-2018. (**Accessed 2025-02-12**)

Ratingen SV, Vonk J, Blokhuis C, Wesseling J, Tielemans E, Weijers E (2021) Seasonal influence on the performance of low-cost NO2 sensor calibrations. Sensors 21(23):7919. https://doi.org/10.3390/s21237919. (**Accessed 2025-02-12**)

Nowack P, Konstantinovskiy L, Gardiner H, Cant J (2021) Machine learning calibration of low-cost NO_2 and PM_{10} sensors: non-linear algorithms and their impact on site transferability. Atmos Measur Tech 14(8):5637–5655. https://doi.org/10.5194/amt-14-5637-2021. (**Accessed 2025-02-12**)

Kim H, Müller M, Henne S, Hüglin C (2022) Long-term behavior and stability of calibration models for NO and NO_2 low cost sensors. Atmos Measur Tech 15(9):2979–2992. https://doi.org/10.5194/amt-2021-433

Malings C, Tanzer R, Hauryliuk A, Kumar SPN, Zimmerman N, Kara LB, Presto AA, Subramanian R (2019) Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. Atmos Measur Tech 12(2):903–920. https://doi.org/10.5194/amt-12-903-2019

Cross ES, Williams LR, Lewis DK, Magoon GR, Onasch TB, Kaminsky ML, Worsnop DR, Jayne JT (2017) Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. Atmos Measur Tech 10(9):3575–3588

Gäbel P, Koller C, Hertig E (2022) Development of Air Quality Boxes Based on Low-Cost Sensor Technology for Ambient Air Quality Monitoring. Sensors 22(10):3830. https://doi.org/10.3390/s22103830. (**Accessed 2024-12-09**)

Cavaliere A, Brilli L, Andreini BP, Carotenuto F, Gioli B, Giordano T, Stefanelli M, Vagnoli C, Zaldei A, Gualtieri G (2023) Development of low-cost air quality stations for next-generation monitoring networks: calibration and validation of NO_2 and O_3 sensors. Atmos Measur Tech 16(20):4723–4740. https://doi.org/10.5194/amt-16-4723-2023. (**Accessed 2025-02-12**)

Levy Zamora M, Buehler C, Datta A, Gentner DR, Koehler K (2023) Identifying optimal co-location calibration periods for low-cost sensors. Atmos Measur Tech 16(1):169–179. https://doi.org/10.5194/amt-16-169-2023

De Vito S, Esposito E, Castell N, Schneider P, Bartonova A (2020) On the robustness of field calibration for smart air quality monitors. Sens Actuators, B Chem 310:127869. https://doi.org/10.1016/j.snb.2020.127869

De Vito S, Esposito E, Massera E, Formisano F, Fattoruso G, Ferlito S, Del Giudice A, D'Elia G, Salvato M, Polichetti T, D'Auria P, Ionescu AM, Di Francia G (2021) Crowdsensing iot architecture for pervasive air quality and exposome monitoring: Design, development, calibration, and long-term validation. Sensors 21(15):5219. https://doi.org/10.3390/s21155219

Fattoruso G, Toscano D, Cornelio A, De Vito S, Murena F, Fabbricino M, Di Francia G (2022) Using mobile monitoring and atmospheric dispersion modeling for capturing high spatial air pollutant variability in cities. Atmosphere 13(11):1933. https://doi.org/10.3390/atmos13111933

De Vito S, Del Giudice A, D'Elia G, Esposito E, Fattoruso G, Ferlito S, Formisano F, Loffredo G, Massera E, Bellucci P, Ciarallo F, Di Francia G (2023) Correlating air pollution concentrations and vehicular emissions in an Italian roadway tunnel by means of low cost sensors. Atmosphere 14(4):679. https://doi.org/10.3390/atmos14040679

Ramel-Delobel M, Peruzzi C, Coudon T, De Vito S, Fattoruso G, Praud D, Fervers B, Salizzoni P (2024) Exposure to airborne particulate matter during commuting using portable sensors: effects of transport modes in a French metropolis study case. J Environ Manage 365:121400. https://doi.org/10.1016/j.jenvman.2024.121400

Bo M, Charvolin-Volta P, Clerico M, Nguyen CV, Pognant F, Soulhac L, Salizzoni P (2020) Urban air quality and meteorology on opposite sides of the alps: The lyon and torino case studies. Urban Clim 34:100698

Beijing Plantower Co., Ltd: PMS7003 Series DataSheet - Digital Universal Particle Concentration Sensor. https://plantower.com/en/products_33/76.html

Alphasense - AMETEK Inc.: CO-A4 Carbon Monoxide Sensor - Datasheet. https://www.alphasense.com/wp-content/uploads/2013/06/COA4.pdf

Alphasense - AMETEK Inc.: NO2-A43F Nitrogen Dioxide Sensor - Datasheet. https://www.alphasense.com/wp-content/uploads/2013/07/NO2A43F.pdf

Alphasense - AMETEK Inc.: OX-A431 Oxidising Gas (O3+NO2) Sensor - Datasheet. https://www.alphasense.com/wp-content/uploads/2013/07/OXA431.pdf

Guangzhou Aosong Electronics Co., Ltd.: AM2302 (DHT22) Temperature and Humidity Sensor Modules. https://cdn.sparkfun.com/assets/f/7/d/9/c/DHT22.pdf

Spinelle L, Gerboles M, Villani MG, Aleixandre M, Bonavitacola F (2015) Field calibration of a cluster of low-cost available sensors for air quality monitoring. part a: Ozone and nitrogen dioxide. Sens Actuators, B Chem 215:249–257

Castell N, Dauge FR, Schneider P, Vogt M, Lerner U, Fishbain B, Broday D, Bartonova A (2017) Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environ Int 99:293–302

Boanini C, Mecca D, Pognant F, Bo M, Clerico M (2021) Integrated mobile laboratory for air pollution assessment Literature review and cc-TrAIRer design. Atmosphere 12(8):1004. https://doi.org/10.3390/atmos12081004

Lewis A, Peltier WR, Schneidemesser E (2018) Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications
Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) Applied linear statistical models
Wood SN (2017) Generalized Additive Models: an Introduction with R. chapman and hall/CRC.
Breiman L (2001) Random forests. Mach Learn 45:5–32
Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
Moore DS, Kirkland S (2007) The Basic Practice of Statistics vol. 2. WH Freeman New York.
Holstius DM, Pillarisetti A, Smith K, Seto E (2014) Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. Atmos Measur Tech 7(4):1121–1131
Mei H, Han P, Wang Y, Zeng N, Liu D, Cai Q, Deng Z, Wang Y, Pan Y, Tang X (2020) Field evaluation of low-cost particulate matter sensors in Beijing. Sensors 20(16):4381
Lopez-Restrepo S, Yarce A, Pinel N, Quintero OL, Segers A, Heemink AW (2021) Urban air quality modeling using low-cost sensor network and data assimilation in the aburrá valley, colombia. Atmosphere 12(1):91
Bisignano A, Carotenuto F, Zaldei A, Giovannini L (2022) Field calibration of a low-cost sensors network to assess traffic-related air pollution along the brenner highway. Atmos Environ 275:119008
Chang JC, Hanna SR (2004) Air quality model performance evaluation. Meteorol Atmos Phys 87(1):167–196
European Parliament and Council: Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. https://eur-lex.europa.eu/eli/dir/2008/50/oj/eng (2008)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Sofia Fellini[1] · Davide Gallione[1] · Vincenzo Vaccaro[1] · Nicole Mastromatteo[1] · Grazia Fattoruso[3] · Marina Clerico[1] · Pietro Salizzoni[1,2]

✉ Sofia Fellini
sofia.fellini@polito.it

[1] Department of Environmental, Land and Infrastructure Engineering, Politecnico di Torino, Corso Duca degli Abruzzi, 10129 Turin, Italy

[2] Laboratoire de Mécanique des Fluides et d'Acoustique, UMR CNRS 5509, Université de Lyon, Ecole Centrale de Lyon, INSA Lyon, Université Claude Bernard Lyon I, 69134 Ecully, France

[3] TERIN-SSI-EDS-LAB, ENEA CR-Portici, 80055 Portici, Italy