# Calibration methodology of low-cost sensors for high-quality monitoring of fine particulate matter

Marie-Laure Aix, Seán Schmitz, Dominique Bicout

# Calibration methodology of low-cost sensors for high-quality monitoring of fine particulate matter

Marie-Laure Aix[a], Seán Schmitz[b], Dominique J. Bicout[c,*]

[a] Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France, marie-laure.aix@univ-grenoble-alpes.fr

[b] Research Institute for Sustainability, Helmholtz Centre Potsdam, Berliner Strasse 130, 14467 Potsdam, Germany, sean.schmitz@rifs-potsdam.de

[c] Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France, dominique.bicout@univ-grenoble-alpes.fr

ABSTRACT.

* Corresponding author at: Laboratoire TIMC, Domaine de la Merci, 38706 La Tronche – France

E-mail address: dominique.bicout@univ-grenoble-alpes.fr

Low concentrations of pollutants may already be associated with significant health effects. An accurate assessment of individual exposure to pollutants therefore requires measuring pollutant concentrations at the finest possible spatial and temporal scales. Low-cost sensors (LCS) of particulate matter (PM) meet this need so well that their use is constantly growing worldwide. However, everyone agrees that LCS must be calibrated before use. Several calibration studies have already been published, but there is not yet a standardized and well-established methodology for PM sensors. In this work, we develop a method combining an adaptation of an approach developed for gas-phase pollutants with a dust event preprocessing to calibrate PM LCS (PMS7003) commonly used in urban environments. From the selection of outliers to model tuning and error estimation, the developed protocol allows to analyze, process and calibrate LCS data using multilinear (MLR) and random forest (RFR) regressions for comparison to a reference instrument. We demonstrate that the calibration performance was very good for $PM_1$ and $PM_{2.5}$ but turns out less good for $PM_{10}$ ($R^2 = 0.94$, RMSE = 0.55 µg/m$^3$, NRMSE = 12% for $PM_1$ with MLR, $R^2 = 0.92$, RMSE = 0.70 µg/m$^3$, NRMSE = 12% for $PM_{2.5}$ with RFR and $R^2 = 0.54$, RMSE = 2.98 µg/m$^3$, NRMSE = 27% for $PM_{10}$ with RFR). Dust events removal significantly improved LCS accuracy for $PM_{2.5}$ (11% increase of $R^2$ and 49% decrease of RMSE) but no significant changes for $PM_1$. Best calibration models included internal relative humidity and temperature for $PM_{2.5}$ and only internal relative humidity for $PM_1$. It turns out that $PM_{10}$ cannot be properly measured and calibrated because of technical limitations of the PMS7003 sensor. This work therefore provides guidelines for PM LCS calibration. This represents a first step towards standardizing calibration protocols and facilitating collaborative research.

KEYWORDS. Air pollution ($PM_1$, $PM_{2.5}$, $PM_{10}$), dust events, sensors calibration, machine learning

## 1. INTRODUCTION.

Air pollution is a leading environmental risk factor, causing around 7 million deaths each year (Fuller et al., 2022). In 2021, new World Health Organization guidelines were established and most of the limits per pollutant were lowered as a result of new evidence from epidemiological studies (World Health Organization, 2021). It was found that even small rises (of 1 $\mu g/m^3$) in pollutants could trigger mortality increases (Danesh Yazdi et al., 2021). Particulate matter (PM) is a complex mixture of solid and liquid particles suspended in air (Adams et al., 2015). In Europe, it is mainly emitted from anthropogenic sources like residential heating, industry, traffic or agriculture (European Environment Agency, 2020). PM represents a threat to the environment and human health (Rai, 2016). Classified as carcinogenic (Loomis et al., 2013), PM induces public health effects depending on its size (Valavanidis et al., 2008). A growing body of research shows that $PM_1$ may be more toxic than $PM_{2.5}$ (Han Wu et al., 2022; Zhang et al., 2020a; Zhang et al., 2020b). A considerate amount of damage is also caused to plants where PM can inhibit photosynthesis and protein synthesis (Rai, 2016). For all these reasons, it is important to effectively measure PM.

Regulatory-grade stations, also called reference monitors, are highly reliable. However, they are expensive (10,000 to 100,000€) and few are implemented, especially in developing countries (deSouza et al., 2020). Moreover, they cannot be easily moved close to people's breathing zone and are not always well located to properly capture population exposure (Duyzer et al., 2015).

Low-cost sensors (LCS) are small portable systems measuring PM at a fine scale, both temporally and spatially. They provide almost real-time data and are affordable, generally less than

€300 per unit. LCS use is increasing worldwide and studies are emerging in developing countries (McFarlane et al., 2021a; Raheja et al., 2022), where official measurement stations are rare. It is expected that LCS will not achieve the same accuracy as reference-grade instruments (Zimmerman, 2022), so it is imperative that LCS undergo a calibration before being used. Once calibrated and quality checked, they compare well with traditional reference instruments and can supplement regulatory-grade networks (Malings et al., 2020).

Various methods are used to calibrate LCS including linear regression (Barkjohn et al., 2021; Hua et al., 2021; Kosmopoulos et al., 2020), multiple linear regression (Barkjohn et al., 2021; Malings et al., 2020; Puttaswamy et al., 2022), mechanistic models based on hygroscopic growth correction (Barkjohn et al., 2021; Crilley et al., 2020; Malings et al., 2020) and machine learning (Kumar and Sahu, 2021; Nowack et al., 2021; Patra et al., 2021). Spatial calibration using regulatory and satellite data also proved to be efficient to calibrate $PM_{2.5}$ (Lu et al., 2021; Mousavi and Wu, 2021). Linear regression (LR) generally lacks accuracy compared to multiple linear regression (MLR) considering several factors (Badura et al., 2019). Mechanistic models accounting for PM aerosol hygroscopic growth can further improve precision. The performance of traditional linear methods is generally limited compared to mechanistic approaches or machine learning. They struggle to capture cross-sensitivities between variables (Jiang et al., 2021; McFarlane et al., 2021a), but their simplicity and transparency are key advantages compared to machine learning.

Although various methodologies have been developed to evaluate LCS performance (Duvall et al., 2021; Fishbain et al., 2017; Languille et al., 2020), there is currently no harmonized protocol to calibrate LCS, which makes calibration studies difficult to start and compare. Calibrating LCS is a complex topic that can involves several questions (Giordano et al., 2021). For example, LCS

users typically wonder how to detect outliers, which calibration model to select, and how long the calibration is valid. In 2019, researchers stated that there was no methodology indicating how to calibrate sensor networks and no ready-to-use open-source code (Barcelo-Ordinas et al., 2019). Since then, scientists developed process workflows (Bi et al., 2020; Chojer et al., 2022; Hong et al., 2021) but none of these studies offered an open-source code or went through a pre-processing step removing dust events. However, it is critical to check the presence of dust events in a dataset (Giordano et al., 2021), especially when using a LCS which can be blind to dust events like PMS7003. Molina Rueda et al. (2023) recently acknowledged that other popular LCS like Sensirion SPS30 and Piera IPS-7100 were also affected by this issue.

Schmitz et al. (2021) tackled complexity and transparency issues by developing a transparent approach to select the best calibration model. Their 7-step method and open-access code were made available for easy implementation (Schmitz et al., 2020). In this study, we employ and adapt this open-source protocol, initially developed for gas-phase pollutants, to calibrate a LCS measuring PM. In addition, we integrate critical measures such as dust events pre-processing and degradation checks to ensure accuracy and reliability. The originality of our approach lies in this holistic methodology which makes a valuable contribution to the field of low-cost PM sensors. Our objective was not to provide calibration formulas for various environments but to develop a complete protocol applicable to different settings. Our objective is not to provide a calibration formula that could be used in various environments but rather to develop a complete protocol applicable to different settings demonstrating hence how it is possible to construct a formula by using the same protocol for different settings. This new framework will be illustrated on a LCS collocation case study conducted in Grenoble.

## 2. MATERIALS AND METHODS.

### 2.1. Field evaluation in Grenoble

#### 2.1.1. Low-cost sensors

Nine low-cost air quality stations measuring $PM_1$, $PM_{2.5}$, $PM_{10}$, temperature, and relative humidity (RH) were assembled. Each air quality station (AQS) had an optical PM sensor (Plantower PMS7003) and a MaxDetect DHT22 sensor measuring internal RH ($RH_{int}$) and temperature ($T_{int}$). PMS7003 usually shows a high correlation with reference instrument (Bauerová et al., 2020; Kang et al., 2021), stability over at least 15 months (Báthory et al., 2021), and good reproducibility (Badura et al., 2019). The AQS operates by drawing PM-carrying air through a pipe (**Figures S1 and S2**). The air then passes through a laser (wavelength ≈ 650 nm), which gets scattered by PM and impacts a photodiode detector (Kelly et al., 2017). The resulting signal is converted into an electrical signal and transmitted to the microcontroller (ESP8266), which uses a proprietary algorithm to calculate PM concentrations in $\mu g/m^3$.

#### 2.1.2. Collocation site, timeframe, and reference device

The study was conducted in Grenoble (France), the largest city in the Alps with 450,000 inhabitants. The local climate is semi-continental with cold winters and hot summers. Before deploying our AQSs, we had to calibrate them through comparison with reference-grade instruments certified for regulatory use. Therefore, we conducted a collocation study with Atmo Auvergne-Rhône-Alpes, the local regulatory instance measuring air quality. We hung our AQSs close to the inlet of their reference monitor (approximately 3m) for 8 months from January 28, 2021 to September 29, 2021. The reference station, called “Les Frênes”, was located close to a park in a low-traffic zone of Grenoble (GPS coordinates: latitude = 45.162°, longitude = 5.735°). Considered as an urban background station, it uses a Fidas Palas GmbH 200, a certified optical

aerosol spectrometer for regulatory air pollution monitoring. Contrary to the LCS, it is equipped with a sampler drawing air at a constant flow rate and an aerosol drying system removing water vapor. The gold standard method to determine PM mass is based on a gravimetric analysis (Giordano et al., 2021; Noble et al., 2001). It involves collecting PM on a pre-weighed filter and subsequently weighing it after sampling and conditioning to eliminate particle-bound water (Wallace and Hopke, 2022). Fidas 200 does not weight particles but uses the light scattered to determine particles sizes and count them. It assumes the particles have a spherical shape and a specific density to calculate their mass concentration using a proprietary algorithm. Fidas 200 is considered equivalent to the gravimetric EN12341 method for urban background environments by the Central Laboratory for Air quality Monitoring (LCSQA, 2017).

#### 2.1.3. Meteorological data

We sourced weather data from ROMMA (Réseau d'Observation Météo du Massif Alpin, 2022), an association maintaining automatic semi-professional stations. The nearest to our collocation site was located 3km away (GPS coordinates: latitude = 45.169°, longitude = 5.768°). A Davis Vantage Pro2 instrument registered all parameters and rainfall (in mm) was measured using an Ultimeter gauge. The following ambient variables were used: temperature ($T_{amb}$, in °C), relative humidity ($RH_{amb}$, in %), and average wind speed (WS, in km/h). WS was measured as a 10-min average, with a measurement frequency of 2.5-3 s. All other data had a 10-min time resolution. A binary time-of-day factor (ToD) from the NOAA Solar Calculator (https://gml.noaa.gov/grad/solcalc/) was extracted to account for diurnal changes. Hourly measurements made during the night were assigned a value of zero, while daytime measurements were given a value of one. All data analyses were performed with RStudio Server 2021.09.1 (Build 372) for Ubuntu Bionic and R statistical software (R Core Team, 2022).

#### 2.1.4. Data cleaning

After collocation, the next crucial step is to prepare data for the first analyses. This involves checking sensors working ranges, performing time alignments to match LCS and reference data, and determining whether to fuse sensor data for comparison. The PMS7003 technical datasheet (Plantower, 2016) specifies a working humidity of 0 to 99%, to which we added a 1% safety margin. Consequently, LCS measurements were filtered out if the internal LCS RH was less than 1% or greater than 98%. Between 0% and 14% of the dataset was removed depending on the AQS. Approximately 78% of RH values exceeding 98% were recorded after 6 months of service. PM concentrations, $RH_{int}$ and $T_{int}$ recorded every 180 s by the LCS but also weather data reported every 10 min, had to be rescaled to hourly values and converted to UTC (Universal Time Coordinated). The value of the current hour represented the average of measurements during the previous hour. As all LCS stations were giving highly correlated results within units over the whole study period (Pearson's correlation coefficient, $r > 0.99$ for $PM_1$, $PM_{2.5}$, $PM_{10}$, $T_{int}$, and $r > 0.97$ for $RH_{int}$), we used the "sensor fusion" method by calculating medians of the nine LCS outputs to ease data processing. Such a clustering method is considered more robust because it reduces calibration errors and inter-sensor variability effects (Barcelo-Ordinas et al., 2019; Smith et al., 2017; Smith et al., 2019).

### 2.2. Target accuracy metrics for PM sensors

Upon completion of the initial data cleaning process, it still remains to check LCS data quality. Checking raw sensor performances to see if calibration is needed should always be the first step. This entails assessing the closeness of the data to the reference values. The US EPA (United States Environmental Protection Agency) developed guidelines for assessing PM sensors (Duvall et al., 2021), including 5 accuracy metrics: Root Mean Square Error (RMSE), Normalized Root Mean

Square Error (NRMSE), coefficient of determination ($R^2$), slope and intercept. A scatter plot is drawn to compare LCS outputs to reference data, and a fitted regression line is used to calculate slope and intercept. If all points align with the bisector, LCS calibration is unnecessary. The US EPA recommends that the intercept be within -5 and +5 µg/m3 with a target slope of 1.0 ± 0.35. $R^2$ should exceed 0,70 for linearity reasons. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}} \quad (1)$$

where $y_i$ and $x_i$ are PM concentrations measured with the reference and LCS, respectively. $R^2$ reflects the accuracy as the closeness between LCS and reference values. For $PM_{2.5,}$ the EPA recommends a RMSE target below 7 µg/m3 and NRMSE ≤ 30%, calculated as:

$$\text{NRMSE} = \frac{\text{RMSE}}{\bar{y}} \times 100 \quad (2)$$

where $\bar{y}$ is the average reference value.

The EPA recommends checking either the RMSE or NRMSE metrics to assess LCS performance. Although the EPA metrics were originally designed for $PM_{2.5}$ performance testing, we also applied them to $PM_1$ and $PM_{10}$. In France, the Central Laboratory for Air quality Monitoring (LCSQA) established uniform calibration criteria for $PM_{2.5}$ and $PM_{10}$, using a relative extended uncertainty (REU) metric rather than RMSE (LCSQA, 2020). The 7 µg/m$^3$ EPA limit for RMSE may not be suitable for $PM_1$ or $PM_{10}$ due to their varying magnitudes. As a result, we mostly used NRMSE when assessing compliance with EPA criteria for $PM_1$ and $PM_{10}$, which also enables easy comparison across studies. We followed the EPA guidelines to analyze our dataset and during the regression analysis of LCS values against the reference as per the US EPA requirement, we observed a distortion of the point cloud (**Figure 1, A**). There were two distinct

patterns for $PM_{2.5}$ and $PM_{10}$ that were associated with two different slopes. Further investigation revealed that the lower slope outliers corresponded to dust events from Sahara Desert.

### 2.3. Pre-processing technique for the detection and removal of dust events

In the European Alps, a significant proportion of dust events comes from the Sahara (Di Mauro et al., 2019). We confirmed that PM measurement was impacted by those events by checking the daily Pearson correlation coefficient (r) between LCS and reference hourly values (**Figure 2, A**). Each time a dust event occurred, r was negative because LCS from Plantower (PMS) can be "blind" to coarse particles like desert dust (Kosmopoulos et al., 2020; Kuula et al., 2020; Vogt et al., 2021). The following two steps method, shown in **Figure 2**, was used on $PM_{2.5}$ to remove dust events on all datasets: (1) select all times with negative r, (2) validate the presence of a dust event with CAMS (Copernicus Atmosphere Monitoring Service) satellite data (retrieved 0.1° x 0.1° resolution dust values from ENSEMBLE dataset (Météo-France, 2020) ('analysis' type)) (**Figure 2, B**). These satellite data were recovered via a free CAMS Web Application Programming Interface (API). The code we developed for extracting dust events from Copernicus API is available in the open-source repository Zenodo (see DOI below). It turned out that all events with negative r corresponded to dust events seen by CAMS. All those events (5% of the full period) were removed from the complete datasets for $PM_1$, $PM_{2.5}$, and $PM_{10}$ (**Figure 2, C**). Next, using analysis of EPA accuracy criteria revealed some metrics did not meet standards, indicating calibration was necessary.

### 2.4. Calibration

After having pre-processed the dataset and checked EPA metrics, we were ready to apply the Schmitz 7-step calibration method. We also intended to compare this method to more traditional approaches that do not use artificial intelligence. Following the same approach as Schmitz et al.,

223 we split the dataset into training and testing sets using a 75/25 ratio for both Schmitz and traditional
224 approaches (**Figure 3**). The first 183 days of the collocation were used to identify the best models,
225 which were then evaluated using the final 60 days.

226 2.4.1. Traditional approaches

227 Linear regression aims at finding a linear relationship between reference measurements
228 (dependent variable) and LCS data (predictor variables). As an example, the equation for $PM_{2.5}$ is:

229 $$PM_{2.5\,ref} = \alpha_0 + \alpha_1 PM_{2.5} \quad (3)$$

230 where $PM_{2.5\ Ref}$ represents reference concentrations, $PM_{2.5}$ the LCS concentrations, $\alpha_1$ the slope
231 of linear regression and $\alpha_0$ the intercept. In the mechanistic approach, temperature and RH are
232 included in the calibration formula. When moisture is high, there is more condensed water on PM
233 surface which scatters more light. This can affect the LCS which tends to give higher values than
234 it should (Wang et al., 2015). As such, numerous studies recommend including RH in calibration
235 equations (Badura et al., 2019; Crilley et al., 2020; Hua et al., 2021). Although with a relatively
236 small influence (Wang et al., 2015), the temperature is also mentioned in various works together
237 with RH as a confounding factor (Chakraborty et al., 2020; Giordano et al., 2021). As an extension
238 of **Eq.(3)**, the following equation involving both $RH_{int}$ and $T_{int}$ is used (Chakraborty et al., 2020;
239 Streibl, 2017):

240 $$PM_{2.5\,ref} = \alpha_0 + \alpha_1 \frac{PM_{2.5lcs}}{g(RH_{int})} + \alpha_2 T_{int} \quad (4)$$

241 where $g(RH_{int})$, the hygroscopic growth factor, defined as the ratio of wet to dry particle
242 diameters at a given RH, is determined by Di Antonio et al. (2018):

243 $$g(RH_{int}) = \left(1 + \kappa \frac{RH_{int}}{100 - RH_{int}}\right)^{\frac{1}{3}} \quad (5)$$

in which κ refers to the degree of hygroscopicity of a particle, depending on the local aerosol. The parameters $\alpha_0$, $\alpha_1$, $\alpha_2$ and $\kappa$ were estimated with the nls() function (R Core Team, 2022).

#### 2.4.2. Schmitz methodology

The Schmitz open-source method is a 7-step methodology (see blue boxes in **Figure 4**) initially developed to calibrate gas-phase pollutants such as nitrogen dioxide ($NO_2$) or ozone ($O_3$). An open-access code can be found online (Schmitz et al., 2020). Step 1 aims to provide a better understanding of the data by analyzing distributions and identifying potential quality issues to be addressed before calibration. This includes checking the distributions of the reference concentrations, LCS data, and weather variables. This is done on both training and testing sets through histograms, violin plots, and time series plots. Step 2 involves removing outliers as they can affect calibration models. Data normality is checked with a Shapiro-Wilk test, followed by a z-test with running mean and standard deviation to detect outliers. Visualization of the outliers on a time-series aids in ensuring a meaningful time window and z-test threshold. Step 3 involves flagging test data falling outside the range of the training data, as well as training data outside the range of the test data. The objective is to identify data points that may be less reliable for prediction and assign them a higher level of uncertainty. Step 4 selects and optimizes the best calibration models, dividing the training dataset into smaller sets using a moving window. In their case-study, Schmitz et al. use a 5 days window, with the models trained on four days and tested on the fifth day. Models with the smaller average RMSE over the various fifth-day predictions are selected. Subsequently, measures of AIC (Akaike Information Criterion) for the MLR model and VI (Variable Importance) for the RFR model are assessed to determine which predictors should remain. VI assesses the relative importance of each predictor variable in predicting the target variable and AIC quantifies the balance between model goodness of fit and complexity. Finally,

267 models are tested on the test subset and assessed using RMSE and $R^2$. The most accurate MLR
268 and RFR models are then sent to the next step. Step 5 allows to validate models by splitting the
269 training set into training and testing subsets at a 75/25 ratio. Those continuous blocks of data allow
270 accounting for autocorrelation in the data. Stability across blocks is then checked using $R^2$, RMSE
271 and VI. Step 6 is about exporting predictions given by the best MLR and RFR models and plotting
272 them using time series. Step 7 evaluates the overall error and confidence intervals associated with
273 the predictions. The reference instrument's technical error is merged with the statistical error
274 associated with the model's predictions. The former is derived from reference specifications, while
275 the latter is computed as the median mean absolute error (MAE) across all blocks during model
276 validation. We adapted this method to our PM dataset (**Figure 5**):
277 (Step 1) Data analysis. Distributions of weather and PM data were similar **(Figure S3**) in both
278 training and testing datasets allowing us to continue the analysis.
279 (Step 2) Outliers removal. Before running the z-test, normality was checked on each sliding
280 window by the detectOutlier function. For $PM_{10}$, between 6.8% and 9% of the data failed the
281 normality test depending on the device. For $PM_{2.5}$, between 7.9% and 9.9% of the data did not pass
282 the normality test, while between 8.6% and 10.8% of $PM_1$ data failed the test. Furthermore,
283 between 2.2% and 3% of the RH data and between 2.6% and 3.3% of the temperature data failed
284 the test. We assumed that most of the data aligned sufficiently with the normal distribution for the
285 z-test to be applicable and we conducted a visual inspection of outliers (**Figure S3**) to ensure that
286 the function did not erroneously flag the extreme values of typical data spikes as outliers. Outliers
287 represented less than 0.1% of the LCS data.
288 (Step 3) Flagging. 12 out-of-bounds points were flagged for $RH_{int}$ (n=10), $T_{int}$ (n=1), and rain
289 (n=1), all occurring during the first two weeks of August (**Figure S3**).

(Step 4) Model selection and tuning.

(a) After conducting an extensive literature search on calibration for LCS measuring PM, we identified key variables and the most promising calibration formulas to be tested (models in **Tables S1, S2, S3 and S4**).

(b) To find the best RFR and MLR models, a moving window of 30 days was used to train models and 10 days to test (Step 4.2 in **Figure 5**). This ratio was chosen proportionally to what was done by Schmitz et al.

(c) The trainControl function (caret R package) was used on all blocks.

(d) The models with the lowest RMSE averaged over the various sliding windows were kept. For RFR, the number of features to be considered at each node of trees (mtry) was adjusted by selecting the best performing in terms of RMSE.

(Step 5) Model validation. RMSE and $R^2$ were considered stable enough (**Figure S4**) compared to what was observed in Schmitz et al.'s work, so we decided to go further.

(Step 6) Model prediction. Final concentrations predictions were exported.

(Step 7) Predictive error calculation (see output example in **Figure S4**). Regarding technical errors in the reference instrument, it was found that the Fidas 200 datasheet (Palas GmbH, 2020) had an uncertainty of 10% for $PM_{2.5}$ and 8% for $PM_{10}$. However, local official comparison tests performed with the gold-standard gravimetric method on the Fidas 200 in Grenoble revealed an expanded uncertainty (k=2) of 11.3% for $PM_{2.5}$ and 13.4% for $PM_{10}$ (LCSQA, 2020) so we applied this expanded uncertainty. We assumed that $PM_1$ had an 11.3% uncertainty like $PM_{2.5}$.

### 2.4.3. Calibrations performances metrics

After having used both traditional and Schmitz approaches to generate calibration formulas, we tested these formulas on the test dataset. To select the best models, we used both $R^2$ (coefficient of determination) and RMSE:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \mathrm{x}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} \quad (6)$$

where $y_i$ are the PM concentrations measured by the reference, $x_i$ represents PM values predicted by the various correction models and $\bar{y}$ is the mean of reference values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}} \quad (7)$$

And after having chosen the best models according to those two criteria, we checked EPA accuracy metrics again. Nonzero intercepts and nonunity slopes are less concerning for optical LCS due to their sensitivity to aerosol characteristics like shape and refractive index. They just have to be calibrated for specific aerosol conditions when accuracy is needed (Giordano et al., 2021; Molina Rueda et al., 2023). $R^2$ indicates the prediction quality of a model and RMSE the accuracy. These two metrics should be given a higher priority to choose the best models. Once the best calibration models have been selected and predictions made, it may be important to ensure the stability of the corrected values over time.

#### 2.4.4. Drift in the corrected measurements over time

An assessment of potential degradation in the corrected measurements over time was conducted using a methodology proposed by deSouza et al. (2023a). To this end, we used PM values from the best calibration formulas (see **Table 1**, formulas in bold) and called them $PM_{Corrected}$. A drift error was then computed for $PM_{10}$, $PM_{2.5}$ and $PM_1$ using this formula:

$$Drift = PM_{Corrected} - PM_{Ref} \quad (8)$$

The goal was to investigate the time-dependence of the error between corrected LCS data and reference values.

## 3. RESULTS.

### 3.1. Raw data analysis

In line with the EPA guidelines, we started our data analysis by drawing scatterplots showing measured LCS values against the reference concentrations with a regression line (**Figure 1, A**). The $R^2$ did not meet EPA compliance standards for $PM_{10}$, and the NRMSE was non-conforming for all PM sizes. During the analysis of the scatterplots, some clusters of points were observed to deviate from the overall trend of the dataset, particularly for $PM_{2.5}$ and $PM_{10}$. These clusters exhibited a smaller slope compared to the majority of the dataset, and further examination revealed that they were associated with Saharan dust events.

### 3.2. Preprocessing with dust events removal

The dust events pre-processing, previously explained in Section 2.3, allowed us to color-code the dust events on the scatterplot (**Figure 1, B**) which visually demonstrated that these outliers would be eliminated with our technique. We then removed all those dust events in the dataset and proceeded with a new quality check according to EPA standards (**Figure 1, C**). Dust events removal yielded notable improvements in $R^2$ values, with a significant 93% increase for $PM_{10}$ and a 23% increase for $PM_{2.5}$. After preprocessing, data from LCS for $PM_1$ and $PM_{2.5}$ already explained 93% and 90% variability of the reference data respectively, whereas $PM_{10}$ had a much lower $R^2$ (0.52), reflecting a poor correlation with REF. NRMSE remained non-compliant for all PM sizes, indicating a need for further calibration. A statistical analysis of the dataset cleared of dust events (**Table S5**) showed that mean $PM_{10}$ reference concentrations were higher than mean LCS

354 concentrations, unlike $PM_{2.5}$ and $PM_1$ where the LCS tended to overestimate the concentrations.
355 ROMMA temperature values were smaller than LCS values, which was not unexpected as the LCS
356 temperature sensor was located inside the device and therefore subject to heating. This might as
357 well have been enhanced by the southern orientation of the collocation site.

358 3.3. Performance of calibration protocols

359 A range of correction formulas (**Table 1**) was generated using the different calibration methods.
360 The best RFR formulas included wind speed for $PM_{2.5}$ and $PM_{10}$. However, the accuracy
361 improvement was not significant compared to formulas solely involving $RH_{int}$ and $T_{int}$. Therefore,
362 we pursued with RF formulas using $RH_{int}$ and $T_{int}$ only. This would make the calibration simpler
363 because LCS systematically provide $RH_{int}$ and $T_{int}$. Moreover, ambient parameters from external
364 weather stations do not represent the local LCS environment, so it is better to use internal LCS
365 parameters for calibration.

366 $PM_1$ models were performing the best in terms of precision and all metrics for $PM_1$ and $PM_{2.5}$,
367 including those obtained through the LR technique, were found to be conform with the EPA
368 checklist. However, the Schmitz protocol outperformed the traditional methods for all PM sizes.
369 Simple LR always performed significantly less well than the other techniques. Compared to LR,
370 mechanistic models brought a 38% accuracy improvement for $PM_1$ and a 13% accuracy
371 improvement for $PM_{2.5}$. MLR worked better for $PM_1$ but RFR performed better (lowest RMSE
372 and highest $R^2$) for $PM_{2.5}$ and $PM_{10}$. None of the methods used to calibrate $PM_{10}$ produced results
373 that met EPA standards. To compare the four model's fitting more in detail, we regressed reference
374 PM concentrations against predicted values on the test dataset (**Figure 6**). It highlights the
375 difficulty in accurately predicting $PM_{10}$ levels, as the scatter in the data is large and $R^2$ does not
376 comply with EPA standards. Corrected $PM_{2.5}$ and $PM_1$ gave more accurate results with a better fit.

### 3.4. $PM_{10}$ calibration improvement essay

To investigate the lower performance of LCS compared to the reference station for $PM_{10}$, we considered the concentration difference between $PM_{10}$ and $PM_{2.5}$ for both reference and LCS as $PM_{(10-2.5)\,Ref} = PM_{10\,Ref} - PM_{2.5\,Ref}$ and $PM_{10-2.5} = PM_{10} - PM_{2.5}$. The reasoning was to try to estimate the concentration of $PM_{2.5}$ counted as $PM_{10}$. We found that the scatterplot of $PM_{(10\text{-}2.5)\,Ref}$ versus $PM_{2.5\,Ref}$ shows almost no relationship with a rather high variability (**Figure S5, A**) whereas $PM_{10\text{-}2.5}$ versus $PM_{2.5}$ exhibits a two-slope linear relation with a breakpoint at 28 $\mu g/m^3$ (**Figure S5, B**). This clearly indicates that although the reference and the LCS both use optical probes, the $PM_{10}$ counting methods (and the transcription algorithms used) lead to qualitatively and quantitatively different results. These two profiles are so different that they cannot be superimposed by making simple transformations such as regressions. The regressions obtained are the best we can do.

### 3.5. The importance of combining preprocessing with the Schmitz method

To evaluate the efficacy of dust events pre-processing, we calibrated our LCS without removing dust events (**Figure 7, C**) and we compared performances with the full combined method (**Figure 7, A**). The RFR and MLR methods used the same predictor variables regardless of whether dust events were present in the dataset or not. Removing dust events was highly beneficial for $PM_{2.5}$, as all metrics exhibited significant improvement. We observed an 11% increase in $R^2$ and a 49% decrease in RMSE for $PM_{2.5}$. The situation was more nuanced for $PM_1$, as dust events removal led to a slight degradation of $R^2$ (-1%) and NRMSE (+3%). For $PM_{10}$, all metrics improved by removing dust events, apart from $R^2$, decreasing by 10%. We also observed that relying solely on dust events processing without using the Schmitz method would not be sufficient to meet the EPA standards (**Figure 7, B**).

### 3.6. Long-term stability of corrected PM measurements

The degradation over the course of the experiment is shown in **Figure S6**. No bias over time was observed on the drift. The mean drift over the whole 8 months period was 0.069 (95% CI: -0.020, 0.157) $\mu g/m^3$ for $PM_{10}$, 0.042 (95% CI: 0.018, 0.066) $\mu g/m^3$ for $PM_{2.5}$ and -0.001 (95% CI: -0.028, 0.026) $\mu g/m^3$ for $PM_1$. Therefore, degradation over time was not an issue in our study.

## 4. DISCUSSION.

Our findings indicate that PMS7003 is more effective in measuring $PM_1$ compared to $PM_{2.5}$, with a lower NRMSE and higher $R^2$. This is consistent with previous research by Molina Rueda et al. (2023) on PMS5003, with similar raw metrics observed. Our results showed an $R^2$ correlation of 0.73 for $PM_{2.5}$ and 0.93 for $PM_1$, which are consistent with their findings of 0.76 and 0.9, respectively. In line with their research, our study also confirms that PMS7003 is a poor probe for measuring $PM_{10}$. First, PMS7003 clearly underestimated increases in $PM_{10}$ as already shown (Sayahi et al., 2019; Vogt et al., 2021). This discrepancy in the $PM_{10}$ between the reference and the LCS is linked to the size distribution of the particles detected as a function of both the optical wavelength of the PMS7003 and the angles of diffusion toward the photodiode. Second, it seems hardly feasible to predict the coarse fraction ($PM_{10-2.5}$) by using the $PM_{2.5}$ cumulative concentrations. The LCS is technically limited to detect $PM_{10}$ so our calibrations could not improve performances tremendously. It is reassuring that models for $PM_1$ and $PM_{2.5}$ are the best performing since these small particles penetrate the respiratory and circulatory system more deeply than $PM_{10}$ and can have stronger toxicity (Valavanidis et al., 2008).

To achieve better alignment with the reference concentrations, four calibration techniques were used to enhance the accuracy of LCS measurements. None of the methods used for $PM_{10}$ produced results that met EPA compliance standards, while all $PM_1$ and $PM_{2.5}$ performance metrics,

including those obtained via the LR technique, met EPA criteria. Among the four tested methods, RFR was the best for $PM_{2.5}$ whereas MLR performed better for $PM_1$. Recently, deSouza et al. (2023b) calibrated a LCS for $PM_{2.5}$ measurement and also stated that RFR gave better results than MLR in a stationary setting. Interestingly, for $PM_1$ we found that a mechanistic approach could perform as well as a bootstrap RFR technique (RMSE = 0.62 µg/m$^3$ for both approaches). But bootstrap MLR still performed better for $PM_1$ (RMSE = 0.55 µg/m$^3$). The best RMSEs found in our study (0.55 µg/m$^3$ for $PM_1$ and 0.70 µg/m$^3$ for $PM_{2.5}$) were excellent compared to what can be seen in the literature. It is difficult to compare RMSEs from one study to another because experimental conditions are different, but we believe that the dust events preprocessing brought a competitive advantage as it is seldom used before calibrating LCS in ambient conditions. We only found a study from Kosmopoulos et al. (2020) removing dust events before calibration, but they only used LR afterwards. Our dust events removal method using satellite analyses could be further automated using CAMS API. In our study, dust events removal immediately increased $R^2$ for $PM_{10}$ and $PM_{2.5}$ but this preprocessing had less influence on $PM_1$. These observations were similar to what had been observed by Kosmopoulos et al. (2020). They stated that the impact of dust events on LCS was more important for $PM_{2.5}$ than for $PM_1$, due to a considerable portion of the smaller-sized particles falling within the 1-2.5 µm range. It would be interesting to develop algorithms to calibrate PM measurements during dust events without discarding them. Stavroulas et al. (2020) used polynomial regression using ratios related to the presence of dust events ($PM_1/PM_{2.5}$ or $PM_{2.5}/PM_{10}$) and had good performances (40% NRMSE improvement) predicting $PM_{2.5}$ with a quadratic equation. Regarding $PM_1$, removing dust events led to a slight degradation of $R^2$ (-1%) and NRMSE (+3%), so we could use the $PM_1$ dataset without discarding dust events. However, for $PM_{2.5}$, dust events removal brought an 11% increase in $R^2$, a 49% decrease in RMSE, and all

other metrics improved as well. Therefore, we recommend checking the presence of dust events by considering scatterplot linearity before step 1 of the Schmitz method. Other types of LCS might react differently to dust events, and the occurrence of dust events may not be consistent across all regions and seasons. However, in this case-study, removing dust events improved considerably the LCS accuracy for $PM_{2.5}$, which was further enhanced by the Schmitz method. Only a few studies had similar accuracy metrics. Puttaswamy et al. (2022) found a 15% NRMSE for $PM_{2.5}$ using LR with temperature, but their study only lasted 100 days. Kelly et al. (2021) had NRMSEs ranging from 13.1% to 22.9% for wildfires using a Gaussian process model. Model accuracy is considered as good when 10% < NRMSE < 20% (Li et al., 2013) and we found NRMSE = 12% for the best $PM_{2.5}$ and $PM_1$ models. As also found by Schmitz et al. on their experimental dataset, models with internal parameters coming from LCS performed better than models using external weather variables. There is an exception yet with WS playing an important role in the first retrieved RFR models for $PM_{2.5}$ and $PM_{10}$. WS was implicated in many LCS studies (Báthory et al., 2021; Hua et al., 2021; Tian et al., 2022). Wind influence seems logical as official measurement stations draw the air inside a pipe at a constant flow rate while the LCS air intake is subject to more variability. It would be interesting to place an anemometer close to the LCS for a more local measure of WS. It seems important to use RH when correcting LCS PM concentrations. Kosmopoulos et al. (2020) found the effects of RH and temperature to be negligible, but their study took place in the Eastern Mediterranean, where RH is smaller than in Grenoble and temperatures are warmer. Many studies concluded that involving RH and temperature can improve models (Badura et al., 2019; Magi et al., 2020). Moreover, RH and temperature might have synergistic effects on sensor performance (Dan Wu et al., 2022). A switch from LR to mechanistic regression involving $RH_{int}$ and $T_{int}$ led to a significant improvement in accuracy (13% for $PM_{2.5}$ and 38% for $PM_1$). This implies that

469 temperature and humidity sensors should work properly. Maintenance on MaxDetect DHT22
470 devices should be considered, especially after 6 months. Another suggestion for improvement
471 would be to replace MaxDetect DHT22 sensor with possibly more robust low-cost sensors like
472 Bosch BME280 or Sensirion SHT85. Even if we did not find any evidence of a significant
473 correlation between rainy days and MaxDetect DHT22 degradation, we could hypothesize that it
474 was enhanced by an unusually humid 2021 summer (Météo-France, 2021). This was flagged
475 during step 3, where out-of-bound points were found during the first two August weeks for $RH_{int}$,
476 $T_{int}$, and rain. In the case study conducted by Schmitz et al. on gas-phase pollutants, the most
477 precise models for predicting $NO_2$ or $O_3$ concentrations on a longer experimental dataset involved
478 $T_{int}$, while $O_3$ models also included $RH_{int}$ and ToD. These findings are consistent with our study,
479 where we observed that internal parameters outperformed ambient parameters. However, ToD had
480 no significant impact on our predictions. In their study, RFR was found to be the better predictor
481 of gas concentrations, albeit only slightly, compared to MLR. Our results showed the same, except
482 for $PM_{10}$, where RFR models performed substantially better than MLR.
483 Degradation over time did not seem to be an issue in this study, in consistency with the 15-month
484 stability reported by Báthory et al. (2021). However, we recommend performing a degradation
485 check because drift might be sensor dependent. It is also important to remind that all these
486 calibration functions remain site-dependent because LCS performance depends on the site-specific
487 PM source mixture (Malyan et al., 2023). We did not perform a validation across different sites,
488 but we had a wide range of weather conditions covered during the 8 months experiment. We do
489 not recommend to blindly transpose these formulas to other sites with different characteristics and
490 aerosols compositions. The applicability of the models stays geographically limited but the
491 Schmitz calibration protocol itself is available online (Schmitz et al., 2020) and can be applied to

any location. We recommend coupling it with the aforementioned preprocessing and drift analysis steps (flowchart in **Figure 4,** see brown boxes). Further research might explore Gaussian Mixture Regression (GMR) to further improve our methodology. This probabilistic technique handles well nonlinear relationships in non-identically distributed explanatory variables. GMR models already demonstrated better $R^2$ and mean absolute error in comparison to MLR and RFR (McFarlane et al., 2021b).

## 5. CONCLUSION.

Taken together, our results show that combining dust events preprocessing with the Schmitz approach can lead to highly accurate results for low-cost sensors measuring PM. Here this combined framework yielded excellent results for $PM_{2.5}$ ($R^2 = 0.92$, RMSE = 0.70 µg/m$^3$, NRMSE = 12%) and performed even better for $PM_1$ ($R^2 = 0.94$, RMSE = 0.55 µg/m$^3$, NRMSE = 12%), although the preprocessing step did not significantly improving $PM_1$ results. $PM_{10}$ could not be properly measured and calibrated due to the technical limitations of the PMS7003. This work represents a novel approach that has not been previously reported in the literature. It demonstrates that without dust events preprocessing, $PM_{2.5}$ data may comply with the EPA guidelines but fall short of achieving optimal accuracy. On the other hand, relying solely on dust events processing without using the Schmitz method would not be sufficient to meet the EPA target metrics. These findings underscore the importance of taking a holistic approach to PM data measurement, which involves integrating multiple techniques to achieve optimal accuracy and compliance with regulatory standards. Moving forward, it may be worthwhile to explore ways to calibrate PM measurements during dust events without discarding them from the dataset, as this could further improve the usefulness of LCS.

## ASSOCIATED CONTENT

### Code availability

The code for the Schmitz methodology can be found in this open-access Zenodo repository: https://doi.org/10.5281/zenodo.4317521 (Schmitz et al., 2020).

The code used to retrieve data from the Copernicus API can be found here: https://doi.org/10.5281/zenodo.7795898 (Aix et al., 2023).

### Supporting information

Air quality station schematic description (Figure S1); Air quality station picture (Figure S2); Seven-step protocol outputs examples (steps 1 to 4) (Figure S3); Seven-step protocol outputs examples (steps 5 to 7) (Figure S4); Scatterplots of $PM_{(10\text{-}2.5)\,Ref}$ versus $PM_{2.5\,Ref}$ and $PM_{10\text{-}2.5}$ versus $PM_{2.5}$ on the train dataset (Figure S5); Drift evaluation: difference between the corrected LCS median measurements (given by the best regression models) and the corresponding hourly reference measurements (Figure S6); list of tested calibration formulas for Multiple Linear Regression using LCS temperature and relative humidity (Table S1); list of tested calibration formulas for Multiple Linear Regression using ambient parameters (Table S2); list of tested calibration equations for Random Forest Regression involving LCS internal temperature and relative humidity (Table S3); list of tested calibration equations for Random Forest Regression using ambient parameters (Table S4); Descriptive statistics for hourly PM concentrations, temperature and relative humidity (Table S5) (PDF).

## AUTHOR INFORMATION

### Authors contributions

Marie-Laure Aix: Conceptualization, Data curation, Formal analysis, Investigation, Software, Visualization, Writing – original draft. Seán Schmitz: Investigation, Methodology, Writing – review, Validation. Dominique J Bicout: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Notes

The authors declare no competing financial interest or personal relationships that could have influenced the work reported in this paper.

## ACKNOWLEDGMENT

Marie-Laure Aix is a PhD student supported by a grant from the Ministry of Education and Research of France through the Ecole Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement (ED-ISCE) of Grenoble Alpes University. This work has been partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003). We would like to warmly thank Atmo Auvergne Rhône-Alpes for their help and assistance in assembling and collocating the AQSs as part of our collaboration. The dust results were generated using the Copernicus Atmosphere Monitoring Service Information [2021]. We also thank ROMMA for the local weather data.

## REFERENCES

Adams, K., Greenbaum, D. S., Shaikh, R., van Erp, A. M., Russell, A. G., 2015. Particulate matter components, sources, and health: Systematic approaches to testing effects. J. Air Waste Manag. Assoc. 65 (5), 544-58. https://doi.org/10.1080/10962247.2014.1001884.

Aix, M. L., Schmitz, S., Bicout, D. J., 2023. Calibration Methodology of Low-Cost Sensors for High-Quality Monitoring of Fine Particulate Matter (1.0.0), Zenodo [code under embargoed Access]. https://doi.org/10.5281/zenodo.7795898.

Badura, M., Batog, P., Drzeniecka-Osiadacz, A., Modzel, P., 2019. Regression methods in the calibration of low-cost sensors for ambient particulate matter measurements. SN Appl. Sci. 1 (6), 622. https://doi.org/10.1007/s42452-019-0630-1.

Barcelo-Ordinas, J. M., Doudou, M., Garcia-Vidal, J., Badache, N., 2019. Self-calibration methods for uncontrolled environments in sensor networks: A reference survey. Ad Hoc Netw. 88, 142-159. https://doi.org/10.1016/j.adhoc.2019.01.008.

Barkjohn, K. K., Gantt, B., Clements, A. L., 2021. Development and Application of a United States wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor. Atmos. Meas. Tech. 4 (6), 4617–4637. https://doi.org/10.5194/amt-14-4617-2021.

Báthory, C., Dobo, Z., Garami, A., Palotas, A., Toth, P., 2021. Low-cost monitoring of atmospheric PM-development and testing. J. Environ. Manage. 304, 114158. https://doi.org/10.1016/j.jenvman.2021.114158.

580 Bauerová, P., Šindelářová, A., Rychlík, Š., Novák, Z., Keder, J., 2020. Low-Cost Air Quality
581 Sensors: One-Year Field Comparative Measurement of Different Gas Sensors and Particle
582 Counters with Reference Monitors at Tušimice Observatory. Atmosphere (Basel). 11 (5),
583 492. https://doi.org/10.3390/atmos11050492.
584
585 Bi, J., Wildani, A., Chang, H. H., Liu, Y., 2020. Incorporating Low-Cost Sensor Measurements
586 into High-Resolution $PM_{2.5}$ Modeling at a Large Spatial Scale. Environ. Sci. Technol. 54
587 (4), 2152-2162. https://doi.org/10.1021/acs.est.9b06046.
588
589 Chakraborty, R., Heydon, J., Mayfield, M., Mihaylova, L., 2020. Indoor Air Pollution from
590 Residential Stoves: Examining the Flooding of Particulate Matter into Homes during Real-
591 World Use. Atmosphere (Basel). 11 (12), 1326. https://doi.org/10.3390/atmos11121326.
592
593 Chojer, H., Branco, P. T. B. S., Martins, F. G., Alvim-Ferraz, M. C. M., Sousa, S. I. V., 2022. Can
594 data reliability of low-cost sensor devices for indoor air particulate matter monitoring be
595 improved? – An approach using machine learning. Atmos. Environ. 286.
596 https://doi.org/10.1016/j.atmosenv.2022.119251.
597
598 Crilley, L. R., Singh, A., Kramer, L. J., Shaw, M. D., Alam, M. S., Apte, J. S., Bloss, W. J.,
599 Hildebrandt Ruiz, L., Fu, P., Fu, W., Gani, S., Gatari, M., Ilyinskaya, E., Lewis, A. C.,
600 Ng'ang'a, D., Sun, Y., Whitty, R. C. W., Yue, S., Young, S., Pope, F. D., 2020. Effect of
601 aerosol composition on the performance of low-cost optical particle counter correction

602 factors. Atmos. Meas. Tech. 13 (3), 1181-1193. https://doi.org/10.5194/amt-13-1181-
603 2020.
604
605 Danesh Yazdi, M., Wang, Y., Di, Q., Requia, W. J., Wei, Y., Shi, L., Sabath, M. B., Dominici, F.,
606 Coull, B., Evans, J. S., Koutrakis, P., Schwartz, J. D., 2021. Long-term effect of exposure
607 to lower concentrations of air pollution on mortality among US Medicare participants and
608 vulnerable subgroups: a doubly-robust approach. Lancet Planet. Health. 5 (10), e689-e697.
609 https://doi.org/10.1016/S2542-5196(21)00204-7.
610
611 deSouza, P., Barkjohn, K., Clements, A., Lee, J., Kahn, R., Crawford, B., Kinney, P., 2023a. An
612 analysis of degradation in low-cost particulate matter sensors. Environ Sci Atmos. 3 (3),
613 521-536. https://doi.org/10.1039/d2ea00142j.
614
615 deSouza, P., Kahn, R. A., Limbacher, J. A., Marais, E. A., Duarte, F., Ratti, C., 2020. Combining
616 low-cost, surface-based aerosol monitors with size-resolved satellite data for air quality
617 applications. Atmos. Meas. Tech. 13 (10), 5319-5334. https://doi.org/10.5194/amt-13-
618 5319-2020.
619
620 deSouza, P., Wang, A., Machida, Y., Duhl, T., Mora, S., Kumar, P., Kahn, R., Ratti, C., Durant, J.
621 L., Hudda, N., 2023b. Evaluating the Performance of Low-Cost $PM_{2.5}$ Sensors in Mobile
622 Settings. arXiv preprint. https://doi.org/10.48550/arXiv.2301.03847.
623

624 Di Antonio, A., Popoola, O. A. M., Ouyang, B., Saffell, J., Jones, R. L., 2018. Developing a
625 Relative Humidity Correction for Low-Cost Sensors Measuring Ambient Particulate
626 Matter. Sensors (Basel). 18 (9), 2790. https://doi.org/10.3390/s18092790.
627
628 Di Mauro, B., Garzonio, R., Rossini, M., Filippa, G., Pogliotti, P., Galvagno, M., Morra di Cella,
629 U., Migliavacca, M., Baccolo, G., Clemenza, M., Delmonte, B., Maggi, V., Dumont, M.,
630 Tuzet, F., Lafaysse, M., Morin, S., Cremonese, E., Colombo, R., 2019. Saharan dust events
631 in the European Alps: role in snowmelt and geochemical characterization. The Cryosphere.
632 13 (4), 1147-1165. https://doi.org/10.5194/tc-13-1147-2019.
633
634 Duvall, R., A. Clements, G. Hagler, A. Kamal, Vasu Kilaru, L. Goodman, S. Frederick, K. Johnson
635 Barkjohn, I. VonWald, D. Greene, Dye, T., 2021. Performance Testing Protocols, Metrics,
636 and Target Values for Fine Particulate Matter Air Sensors: Use in Ambient, Outdoor, Fixed
637 Site, Non-Regulatory Supplemental and Informational Monitoring Applications.
638 EPA/600/R-20/280. U.S. Environmental Protection Agency, Office of Research and
639 Development: Washington, DC, USA.
640 https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=350785&Lab=CEMM
641 (accessed 2023.01.23).
642
643 Duyzer, J., van den Hout, D., Zandveld, P., van Ratingen, S., 2015. Representativeness of air
644 quality monitoring networks. Atmos. Environ. 104, 88-101.
645 https://doi.org/10.1016/j.atmosenv.2014.12.067.

European Environment Agency, 2020. Air Quality in Europe - 2020 Report, EEA Report No 09/2020. https://www.eea.europa.eu/publications/air-quality-in-europe-2020-report (accessed 2023.01.23).

Fishbain, B., Lerner, U., Castell, N., Cole-Hunter, T., Popoola, O., Broday, D. M., Iñiguez, T. M., Nieuwenhuijsen, M., Jovasevic-Stojanovic, M., Topalovic, D., Jones, R. L., Galea, K. S., Etzion, Y., Kizel, F., Golumbic, Y. N., Baram-Tsabari, A., Yacobi, T., Drahler, D., Robinson, J. A., Kocman, D., Horvat, M., Svecova, V., Arpaci, A., Bartonova, A., 2017. An evaluation tool kit of air quality micro-sensing units. Sci. Total. Environ. 575, 639-648. https://doi.org/10.1016/j.scitotenv.2016.09.061.

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., Lichtveld, M., Martin, K., Mustapha, A., Sanchez-Triana, E., Sandilya, K., Schaefli, L., Shaw, J., Seddon, J., Suk, W., Téllez-Rojo, M. M., Yan, C., 2022. Pollution and health: a progress update. Lancet Planet. Health. 6 (6), e535–e547. https://doi.org/10.1016/S2542-5196(22)00090-0.

Giordano, M. R., Malings, C., Pandis, S. N., Presto, A. A., McNeill, V. F., Westervelt, D. M., 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.

Hong, G.-H., Le, T.-C., Tu, J.-W., Wang, C., Chang, S.-C., Yu, J.-Y., Lin, G.-Y., Aggarwal, S. G., Tsai, C.-J., 2021. Long-term evaluation and calibration of three types of low-cost $PM_{2.5}$ sensors at different air quality monitoring stations. Journal of Aerosol Science. 157. https://doi.org/10.1016/j.jaerosci.2021.105829.

Hua, J., Zhang, Y., de Foy, B., Mei, X., Shang, J., Zhang, Y., Sulaymon, I. D., Zhou, D., 2021. Improved $PM_{2.5}$ concentration estimates from low-cost sensors using calibration models categorized by relative humidity. Aerosol Sci. Technol. 55 (5), 600-613. https://doi.org/10.1080/02786826.2021.1873911.

Jiang, Y., Zhu, X., Chen, C., Ge, Y., Wang, W., Zhao, Z., Cai, J., Kan, H., 2021. On-field test and data calibration of a low-cost sensor for fine particles exposure assessment. Ecotoxicol. Environ. Saf. 211, 111958. https://doi.org/10.1016/j.ecoenv.2021.111958.

Kang, Y., Aye, L., Ngo, T. D., Zhou, J., 2021. Performance evaluation of low-cost air quality sensors: A review. Sci. Total. Environ. 818, 151769. https://doi.org/10.1016/j.scitotenv.2021.151769.

689 Kelly, K. E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R., Butterfield,
690 A., 2017. Ambient and laboratory evaluation of a low-cost particulate matter sensor.
691 Environ. Pollut. 221, 491-500. https://doi.org/10.1016/j.envpol.2016.12.039.
692
693 Kelly, K. E., Xing, W. W., Sayahi, T., Mitchell, L., Becnel, T., Gaillardon, P. E., Meyer, M.,
694 Whitaker, R. T., 2021. Community-Based Measurements Reveal Unseen Differences
695 during Air Pollution Episodes. Environ. Sci. Technol. 55 (1), 120-128.
696 https://doi.org/10.1021/acs.est.0c02341.
697
698 Kosmopoulos, G., Salamalikis, V., Pandis, S. N., Yannopoulos, P., Bloutsos, A. A., Kazantzidis,
699 A., 2020. Low-cost sensors for measuring airborne particulate matter: Field evaluation and
700 calibration at a South-Eastern European site. Sci. Total. Environ. 748, 141396.
701 https://doi.org/10.1016/j.scitotenv.2020.141396.
702
703 Kumar, V., Sahu, M., 2021. Evaluation of nine machine learning regression algorithms for
704 calibration of low-cost $PM_{2.5}$ sensor. J. Aerosol Sci. 157, 105809.
705 https://doi.org/10.1016/j.jaerosci.2021.105809.
706
707 Kuula, J., Mäkelä, T., Aurela, M., Teinilä, K., Varjonen, S., González, Ó., Timonen, H., 2020.
708 Laboratory evaluation of particle-size selectivity of optical low-cost particulate matter
709 sensors. Atmos. Meas. Tech. 13 (5), 2413-2423. https://doi.org/10.5194/amt-13-2413-
710 2020.

Languille, B., Gros, V., Bonnaire, N., Pommier, C., Honoré, C., Debert, C., Gauvin, L., Srairi, S., Annesi-Maesano, I., Chaix, B., Zeitouni, K., 2020. A methodology for the characterization of portable sensors for air quality measure with the goal of deployment in citizen science. Sci. Total. Environ. 708, 134698. https://doi.org/10.1016/j.scitotenv.2019.134698.

LCSQA, 2017. Conformité technique des appareils de mesure pour la surveillance des polluants réglementaires – bilan 2016. https://www.lcsqa.org/fr/rapport/2016/ineris-imt-ld/conformite-technique-appareils-mesure-surveillance-polluants-reglementair (accessed 2023.03.19).

LCSQA, 2020. Suivi de l'adéquation des analyseurs automatiques de PM à la méthode de référence : bilan réglementaire 2016-2019 et synthèse des travaux menés depuis 2013. https://www.lcsqa.org/fr/rapport/suivi-de-ladequation-des-analyseurs-automatiques-de-pm-la-methode-de-reference-bilan (accessed 2023.03.19).

Li, M.-F., Tang, X.-P., Wu, W., Liu, H.-B., 2013. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. Energy Convers. Manag. 70, 139-148. https://doi.org/10.1016/j.enconman.2013.03.004.

731 Loomis, D., Grosse, Y., Lauby-Secretan, B., Ghissassi, F. E., Bouvard, V., Benbrahim-Tallaa, L.,
732 Guha, N., Baan, R., Mattock, H., Straif, K., 2013. The carcinogenicity of outdoor air
733 pollution. Lancet Oncol. 14 (13), 1262-1263. https://doi.org/10.1016/S1470-
734 2045(13)70487-X.
735
736 Lu, Y., Giuliano, G., Habre, R., 2021. Estimating hourly $PM_{2.5}$ concentrations at the neighborhood
737 scale using a low-cost air sensor network: A Los Angeles case study. Environ. Res. 195,
738 110653. https://doi.org/10.1016/j.envres.2020.110653.
739
740 Magi, B. I., Cupini, C., Francis, J., Green, M., Hauser, C., 2020. Evaluation of $PM_{2.5}$ measured in
741 an urban setting using a low-cost optical particle counter and a Federal Equivalent Method
742 Beta Attenuation Monitor. Aerosol Sci. Technol. 54 (2), 147-159.
743 https://doi.org/10.1080/02786826.2019.1619915.
744
745 Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A., Subramanian,
746 R., 2020. Fine particle mass monitoring with low-cost sensors: Corrections and long-term
747 performance evaluation. Aerosol Sci. Technol. 54 (2), 160-174.
748 https://doi.org/10.1080/02786826.2019.1623863.
749
750 Malyan, V., Kumar, V., Sahu, M., 2023. Significance of sources and size distribution on
751 calibration of low-cost particle sensors: Evidence from a field sampling campaign. J.
752 Aerosol Sci. 168, 106114. https://doi.org/10.1016/j.jaerosci.2022.106114.

McFarlane, C., Isevulambire, P. K., Lumbuenamo, R. S., Ndinga, A. M. E., Dhammapala, R., Jin, X., McNeill, V. F., Malings, C., Subramanian, R., Westervelt, D. M., 2021a. First Measurements of Ambient $PM_{2.5}$ in Kinshasa, Democratic Republic of Congo and Brazzaville, Republic of Congo Using Field-calibrated Low-cost Sensors. Aerosol Air Qual. Res. 21, 200619. https://doi.org/10.4209/aaqr.200619.

McFarlane, C., Raheja, G., Malings, C., Appoh, E. K. E., Hughes, A. F., Westervelt, D. M., 2021b. Application of Gaussian Mixture Regression for the Correction of Low Cost $PM_{2.5}$ Monitoring Data in Accra, Ghana. ACS Earth Space Chem. 5 (9), 2268-2279. https://doi.org/10.1021/acsearthspacechem.1c00217.

Météo-France, 2021. Un été 2021 assez maussade. https://meteofrance.com/actualites-et-dossiers/actualites/un-ete-2021-assez-maussade (accessed 2023.01.23).

Météo-France, Institut National de l'Environnement Industriel et des Risques (Ineris), Aarhus University, Norwegian Meteorological Institute (MET Norway), Jülich Institut für Energie- und Klimaforschung (IEK), Institute of Environmental Protection – National Research Institute (IEP-NRI), Koninklijk Nederlands Meteorologisch Instituut (KNMI), Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek (TNO), Swedish Meteorological and Hydrological Institute (SMHI), Finnish Meteorological Institute (FMI), 2020. CAMS European air quality forecasts, ENSEMBLE data.

Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store (ADS). https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-europe-air-quality-forecasts?tab=overview (accessed 2023.01.23).

Molina Rueda, E., Carter, E., L'Orange, C., Quinn, C., Volckens, J., 2023. Size-Resolved Field Performance of Low-Cost Sensors for Particulate Matter Air Pollution. Environmental Science & Technology Letters. https://doi.org/10.1021/acs.estlett.3c00030.

Mousavi, A., Wu, J., 2021. Indoor-Generated $PM_{2.5}$ During COVID-19 Shutdowns Across California: Application of the PurpleAir Indoor-Outdoor Low-Cost Sensor Network. Environ. Sci. Technol. 55 (9), 5648-5656. https://doi.org/10.1021/acs.est.0c06937.

Noble, C. A., Vanderpool, R. W., Peters, T. M., McElroy, F. F., Gemmill, D. B., Wiener, R. W., 2001. Federal Reference and Equivalent Methods for Measuring Fine Particulate Matter. Aerosol Science and Technology. 34 (5), 457-464. https://doi.org/10.1080/02786820121582.

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. Meas. Tech. 14 (8), 5637-5655. https://doi.org/10.5194/amt-14-5637-2021.

797 Palas GmbH, 2020. Fidas® 200 S technical datasheet.
798 https://www.palas.de/en/product/download/fidas200s/datasheet/pdf (accessed
799 2023.03.19).
800
801 Patra, S. S., Ramsisaria, R., Du, R., Wu, T., Boor, B. E., 2021. A machine learning field calibration
802 method for improving the performance of low-cost particle sensors. Build. Environ. 190,
803 107457. https://doi.org/10.1016/j.buildenv.2020.107457.
804
805 Plantower, 2016. PMS7003 technical datasheet.
806 https://www.plantower.com/en/products_33/76.html.
807
808 Puttaswamy, N., Sreekanth, V., Pillarisetti, A., Upadhya, A. R., Saidam, S., Veerappan, B.,
809 Mukhopadhyay, K., Sambandam, S., Sutaria, R., Balakrishnan, K., 2022. Indoor and
810 Ambient Air Pollution in Chennai, India during COVID-19 Lockdown: An Affordable
811 Sensors Study. Aerosol Air Qual. Res. 22 (1), 210170.
812 https://doi.org/10.4209/aaqr.210170.
813
814 R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for
815 Statistical Computing, Vienna, Austria. https://www.R-project.org/ (accessed 2023.01.23).
816

817 Raheja, G., Sabi, K., Sonla, H., Gbedjangni, E. K., McFarlane, C. M., Hodoli, C. G., Westervelt,
818 D. M., 2022. A Network of Field-Calibrated Low-Cost Sensor Measurements of $PM_{2.5}$ in
819 Lomé, Togo, Over One to Two Years. ACS Earth Space Chem. 6 (4), 1011−1021.
820 https://doi.org/10.1021/acsearthspacechem.1c00391.
821
822 Rai, P. K., 2016. Impacts of particulate matter pollution on plants: Implications for environmental
823 biomonitoring. Ecotoxicol. Environ. Saf. 129, 120-36.
824 https://doi.org/10.1016/j.ecoenv.2016.03.012.
825
826 Réseau d'Observation Météo du Massif Alpin, 2022. Données Station de Saint-Martin-d'Hères
827 [Members dataset]. https://romma.fr/ (accessed 2023.01.23).
828
829 Sayahi, T., Butterfield, A., Kelly, K. E., 2019. Long-term field evaluation of the Plantower PMS
830 low-cost particulate matter sensors. Environ. Pollut. 245, 932-940.
831 https://doi.org/10.1016/j.envpol.2018.11.065.
832
833 Schmitz, S., Towers, S., Villena, G., Caseiro, A., Wegener, R., Klemp, D., Langer, I., Meier, F.,
834 Von Schneidemesser, E., 2020. Unraveling a black box: An open-source methodology for
835 the field calibration of small air quality sensors (1.0.0), Zenodo [code].
836 https://doi.org/10.5281/zenodo.4317521.
837

838 Schmitz, S., Towers, S., Villena, G., Caseiro, A., Wegener, R., Klemp, D., Langer, I., Meier, F.,
839 Von Schneidemesser, E., 2021. Unravelling a black box: An open-source methodology for
840 the field calibration of small air quality sensors. Atmos. Meas. Tech. 4, 7221–7241.
841 https://doi.org/10.5194/amt-2020-489.
842
843 Smith, K. R., Edwards, P. M., Evans, M. J., Lee, J. D., Shaw, M. D., Squires, F., Wilde, S., Lewis,
844 A. C., 2017. Clustering approaches to improve the performance of low cost air pollution
845 sensors. Faraday Discuss. 200, 621-637. https://doi.org/10.1039/c7fd00020k.
846
847 Smith, K. R., Edwards, P. M., Ivatt, P. D., Lee, J. D., Squires, F., Dai, C., Peltier, R. E., Evans, M.
848 J., Sun, Y., Lewis, A. C., 2019. An improved low-power measurement of ambient $NO_2$ and
849 $O_3$ combining electrochemical sensor clusters and machine learning. Atmos. Meas. Tech.
850 12 (2), 1325-1336. https://doi.org/10.5194/amt-12-1325-2019.
851
852 Stavroulas, I., Grivas, G., Michalopoulos, P., Liakakou, E., Bougiatioti, A., Kalkavouras, P.,
853 Fameli, K., Hatzianastassiou, N., Mihalopoulos, N., Gerasopoulos, E., 2020. Field
854 Evaluation of Low-Cost PM Sensors (Purple Air PA-II) Under Variable Urban Air Quality
855 Conditions, in Greece. Atmosphere. 11 (9). https://doi.org/10.3390/atmos11090926.
856
857 Streibl, N., 2017. Influence of Humidity on the Accuracy of Low-Cost Particulate Matter Sensors.
858 Technical report.
859 https://www.researchgate.net/publication/320474792_Influence_of_Humidity_on_the_A

ccuracy_of_Low-Cost_Particulate_Matter_Sensors?channel=doi&linkId=59e7ad15aca272bc423d0b97&showFulltext=true (accessed 2023.01.23). https://doi.org/10.13140/RG.2.2.21095.75683.

Tian, Y., deSouza, P., Mora, S., Yao, X., Duarte, F., Norford, L. K., Lin, H., Ratti, C., 2022. Evaluating the Meteorological Effects on the Urban Form-Air Quality Relationship Using Mobile Monitoring. Environ. Sci. Technol. 56 (11), 7328-7336. https://doi.org/10.1021/acs.est.1c04854.

Valavanidis, A., Fiotakis, K., Vlachogianni, T., 2008. Airborne particulate matter and human health: toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev. 26 (4), 339-62. https://doi.org/10.1080/10590500802494538.

Vogt, M., Schneider, P., Castell, N., Hamer, P., 2021. Assessment of Low-Cost Particulate Matter Sensor Systems against Optical and Gravimetric Methods in a Field Co-Location in Norway. Atmosphere (Basel). 12 (8), 961. https://doi.org/10.3390/atmos12080961.

Wallace, L., Hopke, P. K., 2022. Measuring Particle Concentrations and Composition in Indoor Air. In: Zhang, Y., Hopke, P. K. and Mandin, C. (Eds.). Handbook of Indoor Air Quality, Singapore: Springer Nature Singapore. pp. 517-567. https://doi.org/10.1007/978-981-16-7680-2_19.

882

883 Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., Biswas, P., 2015. Laboratory Evaluation and
884 Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement.
885 Aerosol Sci. Technol. 49 (11), 1063-1077.
886 https://doi.org/10.1080/02786826.2015.1100710.
887

888 World Health Organization, 2021. WHO global air quality guidelines: particulate matter ($PM_{2.5}$
889 and $PM_{10}$), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.
890 https://apps.who.int/iris/handle/10665/345329 (accessed 2023.01.23).
891

892 Wu, D., Zhang, G., Liu, J., Shen, S., Yang, Z., Pan, Y., Zhao, X., Yang, S., Tian, Y., Zhao, H., Li,
893 J., Cai, L., 2022. Influence of particle properties and environmental factors on the
894 performance of typical particle monitors and low-cost particle sensors in the market of
895 China. Atmos. Environ. 268, 118825. https://doi.org/10.1016/j.atmosenv.2021.118825.
896

897 Wu, H., Zhang, B., Wei, J., Lu, Z., Zhao, M., Liu, W., Bovet, P., Guo, X., Xi, B., 2022. Short-term
898 effects of exposure to ambient $PM_1$, $PM_{2.5}$, and $PM_{10}$ on ischemic and hemorrhagic stroke
899 incidence in Shandong Province, China. Environ. Res. 212 (Pt C), 113350.
900 https://doi.org/10.1016/j.envres.2022.113350.
901

902 Zhang, Y., Ding, Z., Xiang, Q., Wang, W., Huang, L., Mao, F., 2020a. Short-term effects of
903 ambient $PM_1$ and $PM_{2.5}$ air pollution on hospital admission for respiratory diseases: Case-
904 crossover evidence from Shenzhen, China. Int. J. Hyg. Environ. Health. 224, 113418.
905 https://doi.org/10.1016/j.ijheh.2019.11.001.
906
907 Zhang, Y., Fang, J., Mao, F., Ding, Z., Xiang, Q., Wang, W., 2020b. Age- and season-specific
908 effects of ambient particles ($PM_1$, $PM_{2.5}$, and $PM_{10}$) on daily emergency department visits
909 among two Chinese metropolitan populations. Chemosphere. 246, 125723.
910 https://doi.org/10.1016/j.chemosphere.2019.125723.
911
912 Zimmerman, N., 2022. Tutorial: Guidelines for implementing low-cost sensor networks for aerosol
913 monitoring. J. Aerosol Sci. 159, 105872. https://doi.org/10.1016/j.jaerosci.2021.105872.
914
915

LIST OF FIGURES



FIGURE SIZE

Fig. 1: 2-column fitting image.
Fig. 2: 2-column fitting image.
Fig. 3: 1-column fitting image.
Fig. 4: 2-column fitting image.
Fig. 5: 2-column fitting image.
Fig. 6: 2-column fitting image.
Fig. 7: 2-column fitting image.

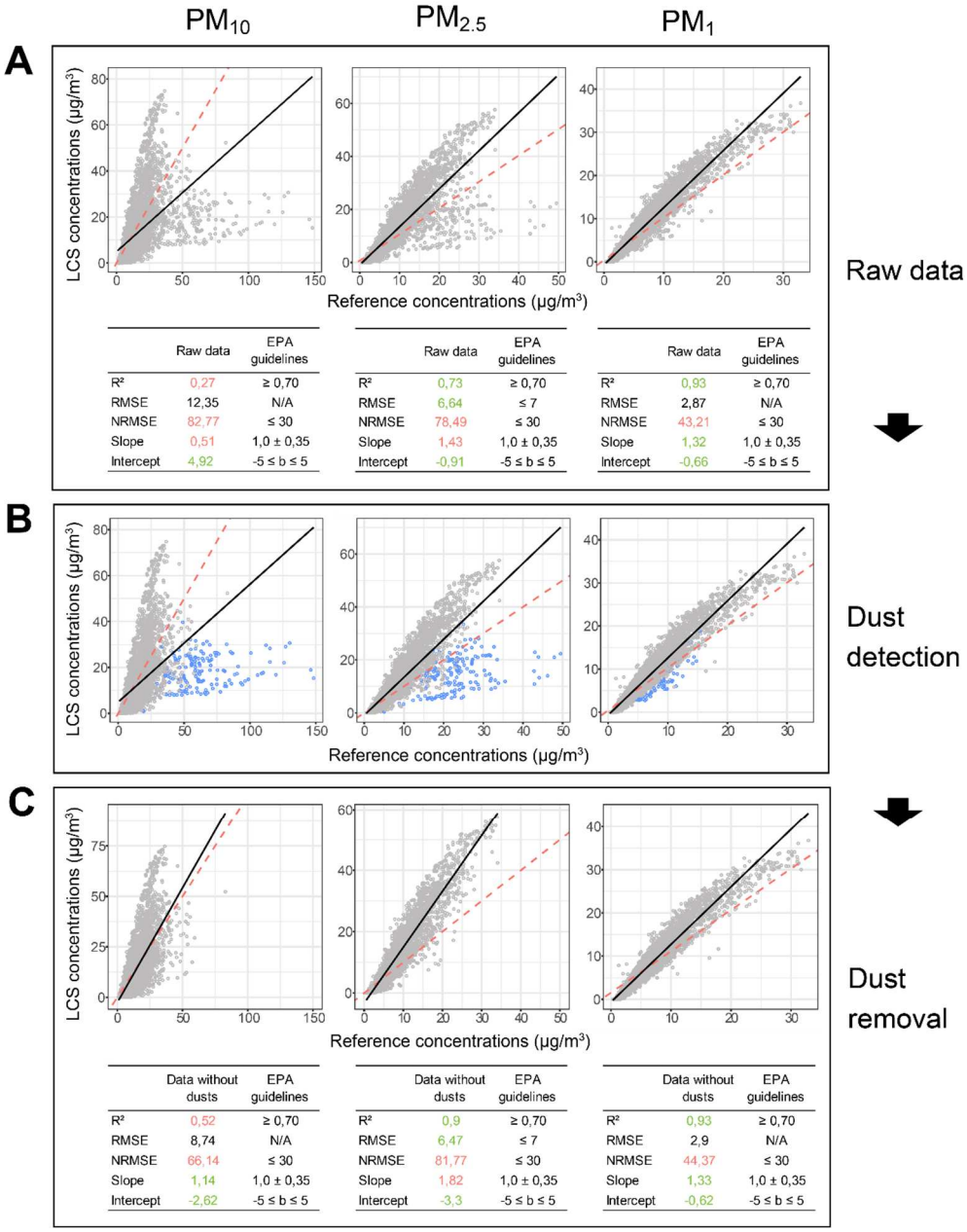
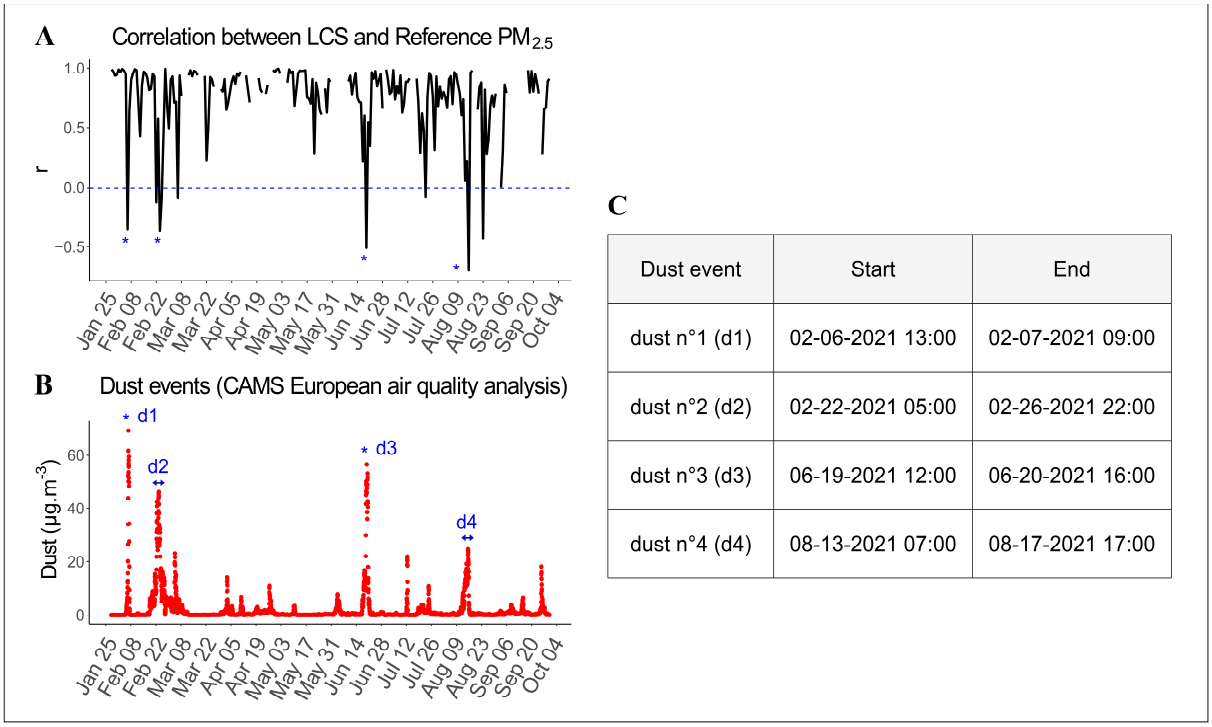



**Figure 1.** Effect of dust pre-processing on regressions of LCS versus reference concentrations (hourly values) and accuracy metrics. The dashed red lines indicate the parity (y = x), and the black lines are the best linear fit. The raw dataset is shown in (A), while (B) displays the dusts identified with our method (in blue), (C) shows the dataset after dust removal. Accuracy metrics such as $R^2$, RMSE (µg/m3), NRMSE (%), intercept (µg/m3) and slope are included (see section 2.2 for metrics definitions). The use of green colour indicates EPA compliant metrics, while red colour denotes non-compliant metrics with EPA standards.

Fig. 1: color




C

| Dust event | Start | End |
|---|---|---|
| dust n°1 (d1) | 02-06-2021 13:00 | 02-07-2021 09:00 |
| dust n°2 (d2) | 02-22-2021 05:00 | 02-26-2021 22:00 |
| dust n°3 (d3) | 06-19-2021 12:00 | 06-20-2021 16:00 |
| dust n°4 (d4) | 08-13-2021 07:00 | 08-17-2021 17:00 |

**Figure 2.** Removal of dust events from the Sahara Desert: (A) Identification of negative daily Pearson correlation coefficients (r), (B) Dust events identification, (C) Selected dust events to remove.

Fig. 2: color


**Figure 3.** Calibration techniques implemented in this study

Fig. 3: color

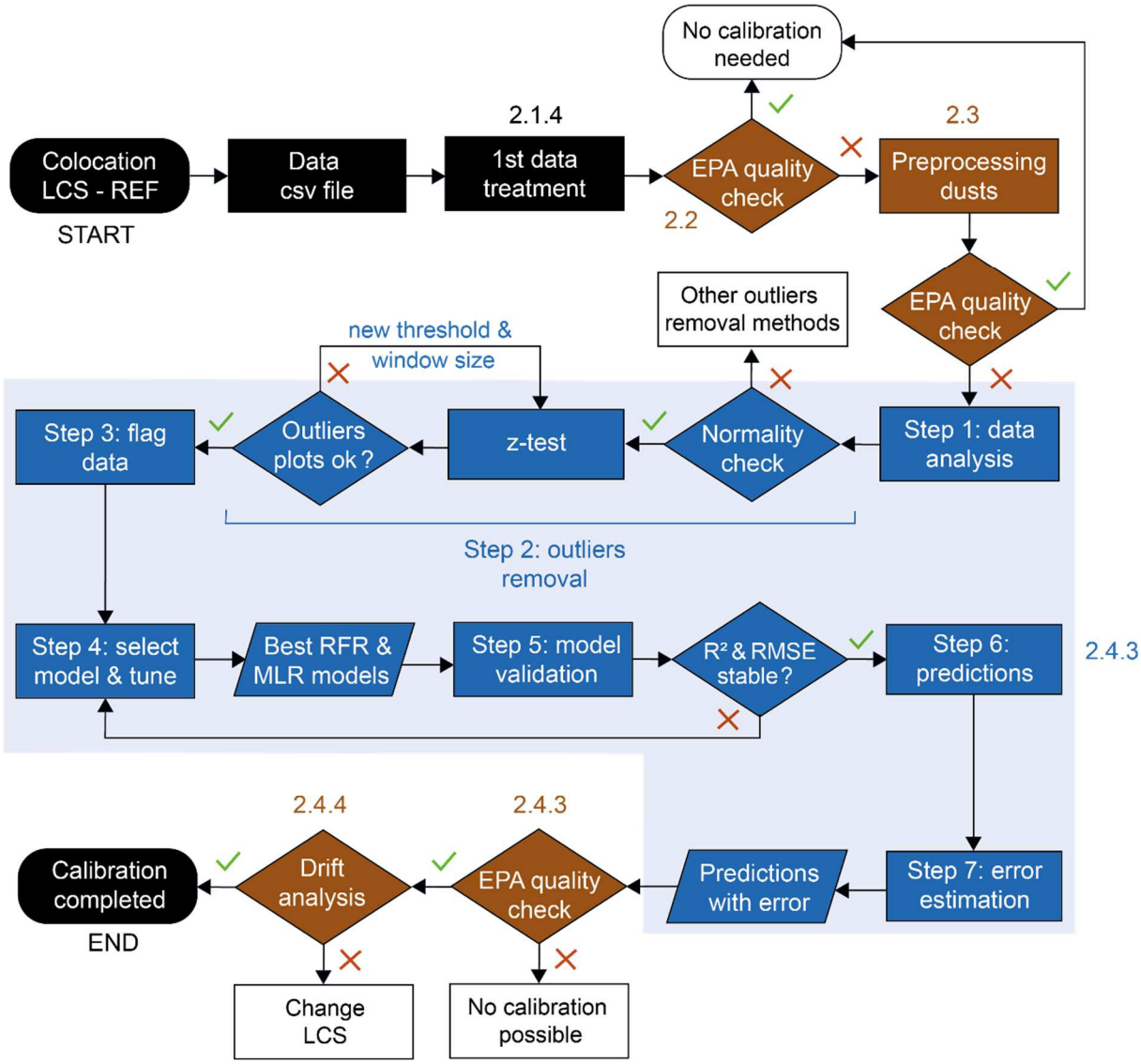



**Figure 4.** Methodology flowchart. In blue, the Schmitz method (section 2.4.2 of the manuscript). In brown, the adaptations made for this PM related study. The numbers (for example 2.1.4) refer to the corresponding Materials and Methods subsections.

Fig. 4: color

STEP 1. RAW DATA ANALYSIS → Variables distribution

STEP 2. OUTLIERS REMOVAL → z-test

| Variables | z-score threshold | Window | Outliers (dataset %) |
|---|---|---|---|
| $PM_1$, $PM_{2.5}$, $PM_{10}$, $T_{int}$, $RH_{int}$ | 6 | 2h | < 0.10 |
| $PM_{1\ ref}$, $PM_{2.5\ ref}$, $PM_{10\ ref}$ | 6 | 10 days | < 0.09 |
| Weather | 6 | 25 days | < 0.25 |

STEP 3. FLAGGING

→ - Flag test data outside train data range
- Flag train data outside test data range

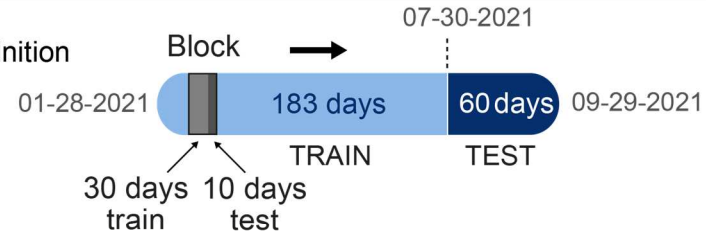
| Variables | Flagged points |
|---|---|
| $T_{int}$ | n = 1 |
| $RH_{int}$ | n = 10 |
| Precipitation | n = 1 |

STEP 4. MODEL SELECTION AND TUNING

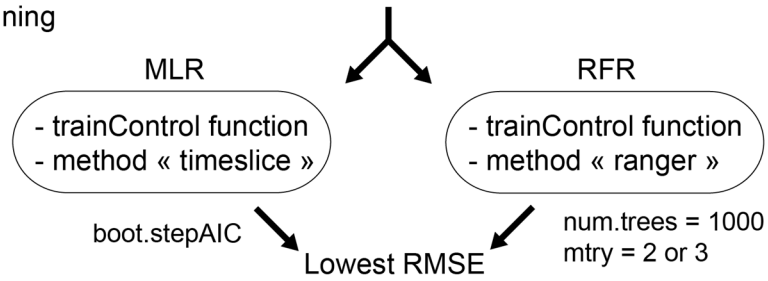
1. Literature → Calibration formulas

$PM_{2.5\ ref} \sim PM_{2.5} + RH_{int} + T_{int}$
$PM_{2.5\ ref} \sim PM_{2.5} + (1/RH_{int}) + T_{int}$
... (Tables S1-S4)

2. Sliding window size definition




3. Model tuning




STEP 5. MODEL VALIDATION → $R^2$ and RMSE stability across blocks

STEP 6. MODEL PREDICTIONS → Export concentrations predictions

STEP 7. PREDICTIVE ERROR CALCULATION

= Reference technical error ($PM_1$, $PM_{2.5}$ : 11.3%, $PM_{10}$ : 13.4%) + Statistical error (Mean absolute error)

**Figure 5.** Flowchart of adapting the Schmitz method

Fig. 5: color

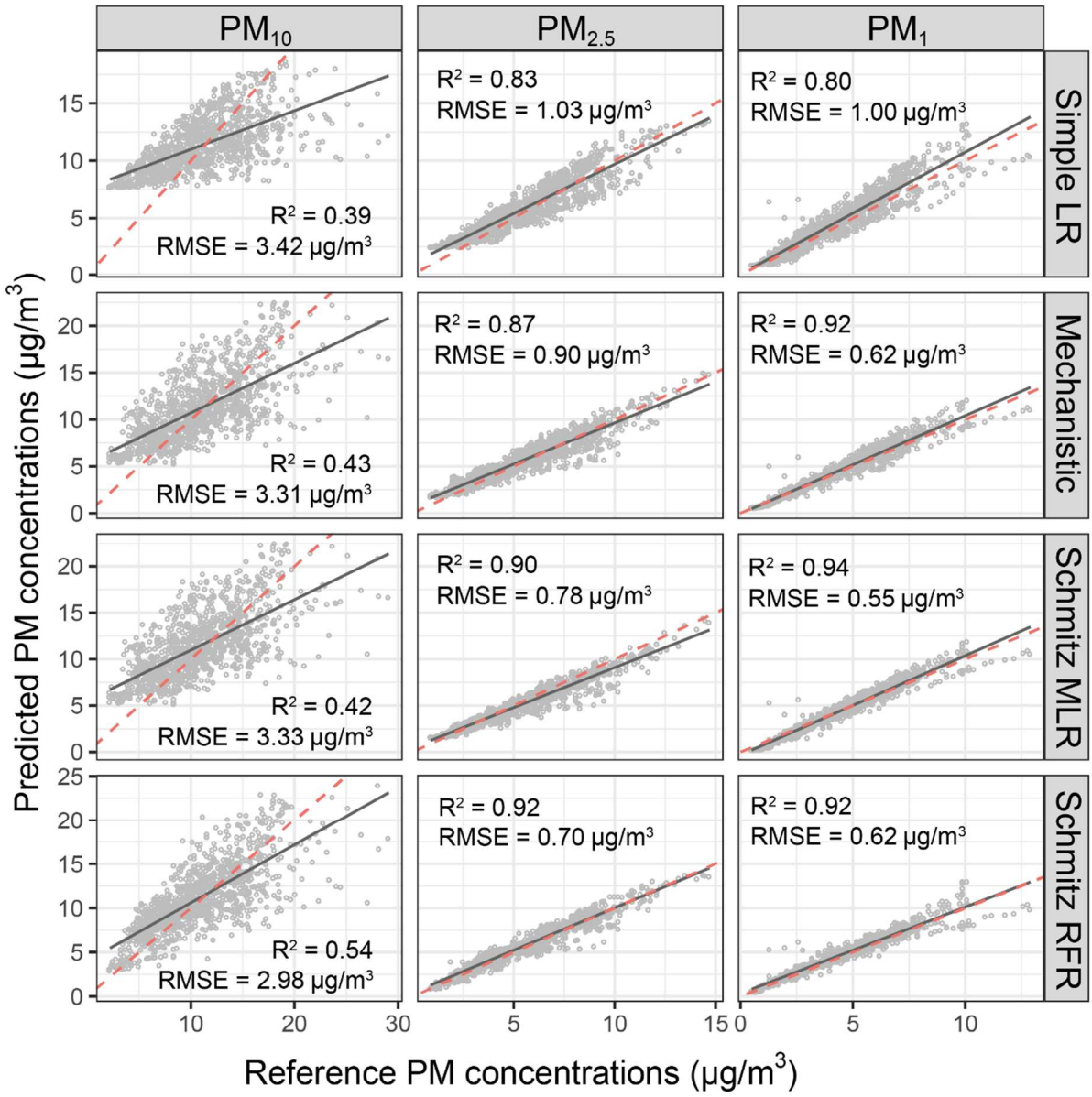



**Figure 6.** Scatterplot of hourly PM reference concentrations versus predicted PM concentrations (test dataset). The red dashed line shows the parity line (y = x) and the black solid line shows the fitted linear regressions.

Fig. 6: color

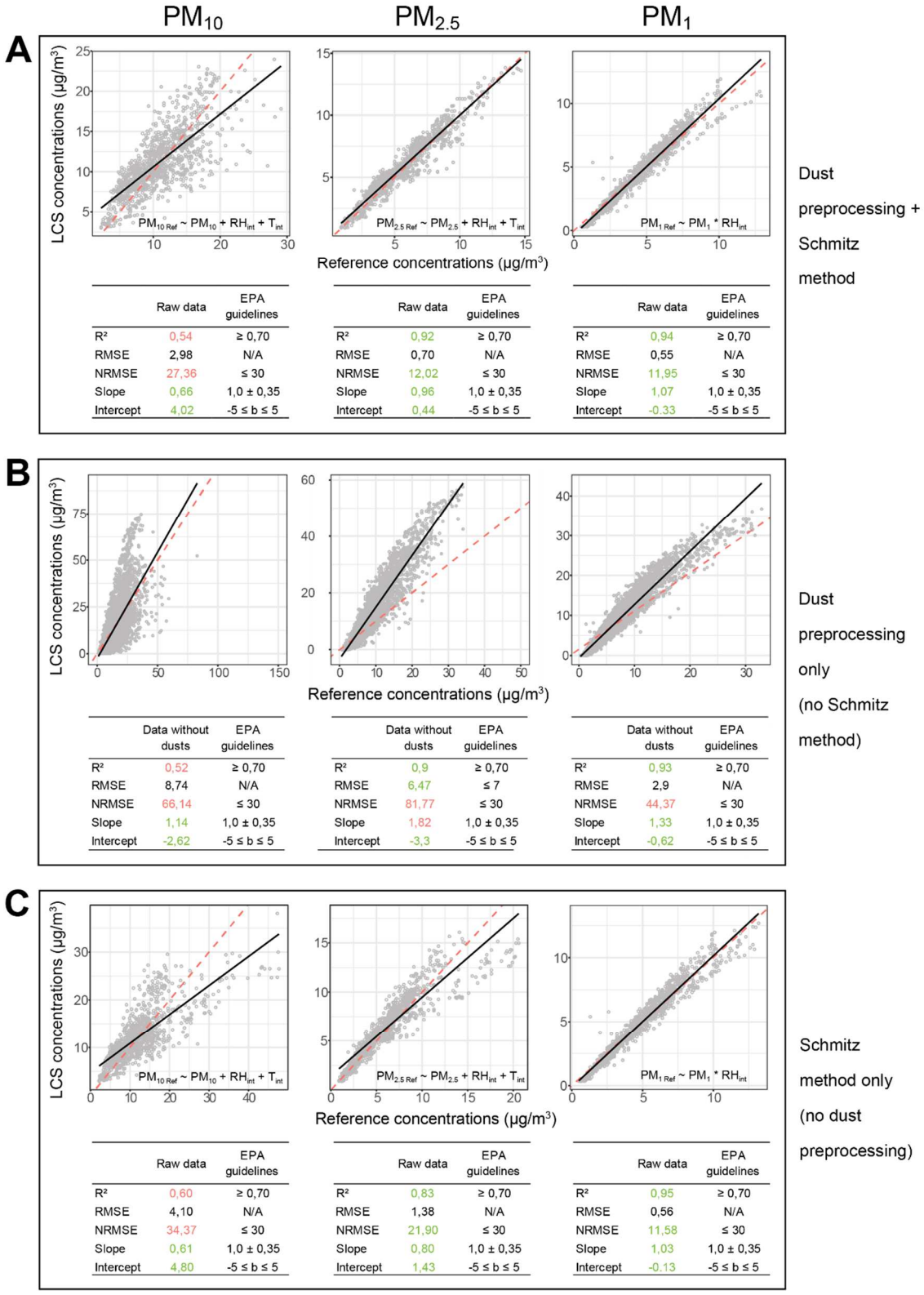



**Figure 7.** Calibration performances: (A) complete method involving the Schmitz technique and dust event preprocessing, (B) with removal of dust events but without having yet applied the Schmitz method, (C) with the Schmitz method but without dust event removal. The green color indicates EPA compliant metrics, while red values denote non-EPA compliant metrics.

Fig. 7: color

**Table 1.** Model performances on the test dataset, using hourly values.

| Method | Correction formula | $R^2$ | RMSE[a] | NRMSE[b] | Slope | Intercept |
|---|---|---|---|---|---|---|
| | **$PM_1$** | | | | | |
| LR | $PM_{1\ Ref} = 0.70\ PM_1 + 0.93$ | 0.80 | 1.00 | 21.70 | 1.06 | 0.12 |
| Mechanistic | $PM_{1\ Ref} = 0.08 + 0.83 \frac{PM_1}{\left(1 + 0.13 \frac{RH_{int}}{100 - RH_{int}}\right)^{\frac{1}{3}}} + 0.03\ T_{int}$ | 0.92 | 0.62 | 13.42 | 1.05 | -0.05 |
| **MLR** | **$PM_{1\ Ref} = PM_1 * RH_{int}$** | **0.94** | **0.55** | **11.95** | **1.07** | **-0.33** |
| RFR | $PM_{1\ Ref} = PM_1 + RH_{int} + T_{int}$ | 0.92 | 0.62 | 13.45 | 0.98 | 0.33 |
| | **$PM_{2.5}$** | | | | | |
| LR | $PM_{2.5\ Ref} = 0.49\ PM_{2.5} + 2.49$ | 0.83 | 1.03 | 17.75 | 0.86 | 1.12 |
| Mechanistic | $PM_{2.5\ Ref} = 0.50 + 0.65 \frac{PM_{2.5}}{\left(1 + 0{,}25 \frac{RH_{int}}{100 - RH_{int}}\right)^{\frac{1}{3}}} + 0.08\ T_{int}$ | 0.87 | 0.90 | 15.57 | 0.88 | 0.84 |
| MLR | $PM_{2.5\ Ref} = PM_{2.5} * RH_{int}$ | 0.90 | 0.78 | 13.46 | 0.86 | 0.48 |
| RFR | $PM_{2.5\ Ref} = PM_{2.5} + RH_{int} + T_{int} + WS$ | 0.92 | 0.70 | 12.02 | 0.96 | 0.46 |
| | **$PM_{2.5\ Ref} = PM_{2.5} + RH_{int} + T_{int}$** | **0.92** | **0.70** | **12.02** | **0.96** | **0.44** |
| | **$PM_{10}$** | | | | | |
| LR | $PM_{10\ Ref} = 0.45\ PM_{10} + 7.71$ | 0.39 | 3.42 | 31.45 | 0.33 | 7.63 |
| Mechanistic | $PM_{10\ Ref} = 1.75 + 1.12 \frac{PM_{10}}{\left(1 + 2{,}68 \frac{RH_{int}}{100 - RH_{int}}\right)^{\frac{1}{3}}} + 0{,}23\ T_{int}$ | 0.43 | 3.31 | 30.40 | 0.53 | 5.40 |
| MLR | $PM_{10\ Ref} = 1.15 + 0.89 \frac{PM_{10}}{\left(1 + \frac{RH_{int}}{100 - RH_{int}}\right)^{\frac{1}{3}}} + 0{,}27\ T_{int}$ | 0.42 | 3.33 | 30.61 | 0.54 | 5.53 |
| RFR | $PM_{10\ Ref} = PM_{10} + RH_{int} + T_{int} + WS$ | 0.54 | 2.96 | 27.21 | 0.64 | 4.29 |
| | **$PM_{10\ Ref} = PM_{10} + RH_{int} + T_{int}$** | **0.54** | **2.98** | **27.36** | **0.66** | **4.02** |

$PM_{1\ Ref}$ = $PM_1$ concentrations in µg/m$^3$ measured by the reference, $PM_1$ = $PM_1$ concentrations in µg/m$^3$ measured by the LCS, $RH_{int}$ = internal relative humidity in % measured by the LCS, $T_{int}$ = internal (LCS) temperature in °C, WS = average wind speed in km/h, * = the model has interactions. [a] in µg/m$^3$. [b] NRMSE or normalized RMSE (in %) = RMSE divided by the mean of the observed (reference) concentrations and multiplied by 100.

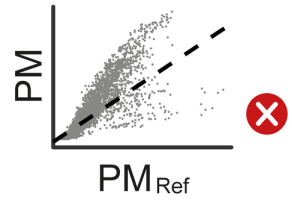
## COLOCATION

PM, $T_{int}$, $RH_{int}$

LCS

$PM_{Ref}$

REF



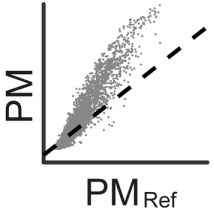

## CALIBRATION

**Preprocessing dusts**

+

**Seven steps method**

1. Data analysis
2. Outliers
3. Flagging
4. Model tuning
5. Validation
6. Predictions
7. Error




## OUTPUT