sensors

MDPI

Article

An Investigation on the Possible Application Areas of Low-Cost PM Sensors for Air Quality Monitoring

Domenico Suriano * and **Mario Prato**

ENEA—Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Department for Sustainability—Brindisi Research Center, SS. 7 Appia, km 706, 72100 Brindisi, Italy
* Correspondence: domenico.suriano@enea.it

Abstract: In recent years, the availability on the market of low-cost sensors (LCSs) and low-cost monitors (LCMs) for air quality monitoring has attracted the interest of scientists, communities, and professionals. Although the scientific community has raised concerns about their data quality, they are still considered a possible alternative to regulatory monitoring stations due to their cheapness, compactness, and lack of maintenance costs. Several studies have performed independent evaluations to investigate their performance, but a comparison of the results is difficult due to the different test conditions and metrics adopted. The U.S. Environmental Protection Agency (EPA) tried to provide a tool for assessing the possible uses of LCSs or LCMs by publishing guidelines to assign suitable application areas for each of them on the basis of the mean normalized bias (MNB) and coefficient of variance (CV) indicators. Until today, very few studies have analyzed LCS performance by referring to the EPA guidelines. This research aimed to understand the performance and the possible application areas of two PM sensor models (PMS5003 and SPS30) on the basis of the EPA guidelines. We computed the R^2, RMSE, MAE, MNB, CV, and other performance indicators and found that the coefficient of determination (R^2) ranged from 0.55 to 0.61, while the root mean squared error (RMSE) ranged from 11.02 $\mu g/m^3$ to 12.09 $\mu g/m^3$. Moreover, the application of a correction factor to include the humidity effect produced an improvement in the performance of the PMS5003 sensor models. We also found that, based on the MNB and CV values, the EPA guidelines assigned the SPS30 sensors to the "informal information about the presence of the pollutant" application area (Tier I), while PMS5003 sensors were assigned to the "supplemental monitoring of regulatory networks" area (Tier III). Although the usefulness of the EPA guidelines is acknowledged, it appears that improvements are necessary to increase their effectiveness.

Keywords: air quality monitoring; low-cost sensors; EPA guidelines; PM sensors; sensor evaluation; field evaluation; air pollutants; gravimetric method

check for updates

Citation: Suriano, D.; Prato, M. An Investigation on the Possible Application Areas of Low-Cost PM Sensors for Air Quality Monitoring. *Sensors* **2023**, *23*, 3976. https://doi.org/10.3390/s23083976

Academic Editor: Manuel Aleixandre

Received: 27 February 2023
Revised: 30 March 2023
Accepted: 11 April 2023
Published: 14 April 2023

1. Introduction

Many studies have proved the existence of a direct link between exposure to air pollutants and issues concerning public health or climate change [1–5]. Air quality monitoring is controlled by national regulations, and the equipment required to meet the standards established by such regulations is characterized by high costs due to purchasing, maintenance, and logistical issues [6–11]. For this reason, in many cases, fixed monitoring station networks of governmental agencies feature few nodes that are sparsely deployed across the territory. As a consequence, it is not often possible to obtain pollutant maps with an adequate spatio-temporal resolution [9,12].

In recent years, an appealing solution to this issue has been represented by the rising of air quality monitors based on low-cost sensors [10–12]. A remarkable number of research institutions and companies have started to design, produce, and test a huge variety of sensors not only for pollutant gas and particulate matter monitoring [10,11,13], but also for malodor detection [14,15]. The use of low-cost sensors (LCSs) or low-cost monitors

(LCMs) based on LCSs for air quality monitoring has been investigated and explored by several studies reporting interesting potentialities, but also substantial limitations and caveats [9–13,16–19].

The technologies featuring LCSs and LCMs provide devices that are ten or more times cheaper than the regulatory instrumentation [9,11], but their data quality is questionable [9,11,17,19]; additionally, the performance information provided by the manufacturers of LCMs/LCSs is limited in most cases. Several studies have already addressed this issue, and various strategies have been explored to improve LCS or LCM performance. These range from the employment of sensor arrays [20] to the use of various data elaboration algorithms, such as multilinear regression or artificial neural networks [9,11,20,21]. The process of improving the performance of these devices by post-processing their data is commonly termed calibration.

Several studies suggest that it is preferable to evaluate or calibrate the performance of such devices in the environment of their final deployment (more concisely, "on-field"), which could be an outdoor site or an indoor space [9,11,13,16,18,19,21]. The on-field evaluation or calibration of LCSs or LCMs is performed by co-locating the device under test with reference instrumentation featuring higher standards of accuracy and precision [11].

The performance of the devices under test can be assessed through indicators calculated utilizing the data provided by the reference instrumentation and the data of the devices under evaluation or calibration. We found that the most commonly used indicators in studies concerning LCS/LCM evaluation or calibration were the coefficient of determination (R^2), the root mean squared error (RMSE), the mean absolute error (MAE), the mean normalized bias (MNB), and the coefficient of variation (CV). The R^2 indicator describes how well the LCM/LCS correlates with the reference device; it ranges from 0 to 1. Values close to 0 indicate poor performance, while values close to 1 show a good agreement between the device under test and the reference device. RMSE, MAE, and MNB are indicators related to the extent of the error between the measurements of the LCM/LCS and the measurements of the reference device; values close to 0 represent a good performance. The coefficient of variation (CV) is used to describe the extent of the variation displayed in the measurements provided by several samples of the same LCM/LCS model under evaluation; values close to 0 indicate a good level of consistency for the model.

The study described in this manuscript focused on the evaluation of the performance of two LCS models designed for measuring particulate matter (PM) concentrations and their application areas. PM is an air pollutant composed of microscopic particles whose aerodynamic diameter is less than or equal to 10 μm, in the case of PM_{10}, or less than or equal to 2.5 μm or 1 μm, in the case of $PM_{2.5}$ and PM_1, respectively.

The present manuscript is organized as follows. In Section 2, an overview of previous related works is presented together with an overall description of this study. Section 3 provides some necessary background information that is useful to fully understand some fundamental aspects of this research. The materials and methods used to perform this study are reported in Section 4, while the results are shown in Section 5. A detailed discussion of the results can be found in Section 6. The findings presented in this penultimate section led to the conclusions summarized in Section 7 of this article.

2. Related Works and Study Description

Although in recent years, several studies concerning the evaluation/calibration of PM sensors already available on the market have been conducted, it is quite difficult to compare their results due to the remarkable heterogeneity of conditions under which they were performed. By reading the scientific literature, it was found that they differed in terms of the test environment (outdoor, indoor, or in a laboratory test chamber), reference instrumentation used, performance indicators, dataset structures (e.g., data grouped by hourly or daily means), and test duration. All these aforementioned factors directly affected the quantification of the performance, and this element was the origin of the difficulties in comparing the results, even though we considered studies that used identical indicators.

Gao et. al. [22] evaluated the performance of a Shinyei sensor measuring $PM_{2.5}$ during a 4-day test performed in an outdoor environment, concluding that it correlated better with optical reference instruments (R^2 = 0.86–0.89) than with gravimetric ones (R^2 = 0.53). Vogt, Castell, et al. [23] carried out a performance evaluation of the PMS5003, SPS30, and OPC-N3 sensors through an outdoor test that lasted 7 weeks by using optical and gravimetric reference instruments. They found that in the case of $PM_{2.5}$, the sensors showed a good performance in terms of the coefficient of determination (R^2 = 0.7–0.9), while they performed worse for PM_{10} measurements (R^2 = 0.6–0.7). Kosmopoulos, Kazantzidis, et al. [24] evaluated and calibrated a PMS5003 sensor integrated into an LCM called PurpleAir-PA II by performing an experiment that lasted 18 months. In this experiment, the authors gathered data representing hourly measurements provided by the LCM and an optical counter as a reference, reporting R^2 = 0.81 for PM_1, R^2 = 0.56 for $PM_{2.5}$, and $R^2 < 0.37$ for PM_{10}. Masic et al. [25] evaluated the PMS5003 and OPC-N2 sensors through a test carried out in a heavily polluted outdoor environment. They showed that by considering daily averaged measurements, the sensors correlated well with the reference (R^2 = 0.9–0.95), even though the MAE ranged between 29.4 $\mu g/m^3$ and 55.2 $\mu g/m^3$.

Other works [26–28] have explored the potentialities and limits of various PM sensor models by applying them in different scenarios.

Evaluations of a notable variety of LCSs/LCMs for PM concentration measurements can be found in the AQ-SPEC program [18]. This study has remarkable importance not only for the number of different LCM/LCS models investigated, but also because they were all evaluated under the same conditions. In this work, the devices under test were compared on-field or in the laboratory with different types of reference device: beta attenuation monitors and optical counters. The duration of the evaluation period was fixed at roughly two months for every device tested, and the data gathered were grouped into 5 min, hourly, and daily means. The indicators used for the evaluations were the coefficient of determination (R^2) and the coefficient of variation (CV).

The study presented in this manuscript was performed under similar conditions: two copies of two different PM sensor models (PMS5003, produced by Plantower, and SPS30, produced by Sensirion [29,30]) were evaluated on-field through a test that lasted roughly two months. The purpose of our study was to understand the potential uses, or application areas, of these PM sensors presenting a good price–quality ratio. To accomplish this task, we took as a reference the Williams et al. [19] report published by the U.S. Environmental Protection Agency (EPA). In this report, the potential uses of LCSs were classified on the basis of certain performance indicators, more precisely, the MNB and CV [13,19]. In particular, we investigated if the PM sensors could be used for supplemental monitoring to complement the existing fixed PM stations of the local government environmental protection agency named ARPA Puglia [31]. As indicated in the EPA report [19], if the sensor performance presents a variance and a bias error of less than 20% (CV and MNB) and a data completeness greater than 80%, the devices can be reliably used as a supplemental monitoring tool for improving the spatial resolution of the pollutant concentration maps produced by ARPA Puglia [32]. As seen in previous works [18,23], the reference instrumentation used for evaluating the LCSs/LCMs directly affects the indicator performance; moreover, SPS30 and PMS5003 have already been evaluated in the AQ-SPEC program [18] using beta attenuation monitors and optical counters as a reference. Considering these factors, we judged it important to use the gravimetric measurements provided by ARPA Puglia in order to understand the possibility of employing these PM sensors to complement ARPA Puglia data. These considerations, in addition to the fact that the MNB of the sensors was not calculated in the AQ-SPEC project [18], led to the design of the experiment hereafter described.

3. Background

As mentioned earlier, the conditions characterizing the assessment or the calibration of the LCSs/LCMs are very important for their final evaluation. The test environment is the most influential factor. In general, LCMs or LCSs evaluated or calibrated in a laboratory test chamber show better performance in comparison with devices evaluated or calibrated on-field. This is because the environmental variables typical of real-world scenarios, which negatively affect LCM/LCS performance, are hard to reproduce in a laboratory test chamber. Other relevant factors weighing on their performance quantification are the type of data grouping, the instrumentation used as a reference, the duration of the evaluation period, the range of the pollutant concentrations, and the performance indicators adopted. As already mentioned, there is no commonly accepted procedure used to perform LCM/LCS evaluations or calibrations. In this respect, the EPA guidelines tried to standardize these processes by proposing a minimum set of rules to allow a better comparison of LCS/LCM performance. In particular, they proposed the adoption of the CV and MNB indicators to provide reliable information about the optimal use of LCSs/LCMs.

Therefore, in order to assess the possible uses of the sensors following the approach indicated in the EPA report [19], the MNB and CV indicators were calculated for each sensor model. In addition to these, the coefficient of determination (R^2), RMSE, and MAE were also computed for each copy of the sensors to allow a comparison with previous studies where the same sensors were evaluated or calibrated. R^2, MAE, MNB, RMSE, and CV are performance indicators defined as follows (see also [9,13,21,23,33]):

$$R^2 = \frac{\left(\sum_{i=1}^{N}\left(s_i - \bar{s}\right)\left(r_i - \bar{r}\right)\right)^2}{\sum_{i=1}^{N}\left(s_i - \bar{s}\right)^2\left(r_i - \bar{r}\right)^2} \tag{1}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|s_i - r_i| \tag{2}$$

$$MNB = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N}\frac{(s_{i,j} - r_i)}{r_i} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - r_i)^2} \tag{4}$$

$$CV = \frac{\sqrt{\frac{1}{MN-1}\sum_{j=1}^{M}\sum_{i=1}^{N}\left(s_{i,j} - s_i\right)^2}}{\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\sum_{i=1}^{N}s_{i,j}} \tag{5}$$

where r_i is the ith measurement of the reference, s_i is the ith reading of the sensor, N is the total number of observations, $\bar{s}$ is the average of the sensor readings, and $\bar{r}$ is the average of the reference measurements. Concerning the CV and MNB formulas, $s_{i,j}$ represents the ith reading of the sensor of the jth copy of the sensor model, while M is the number of copies for each sensor model, which was equal to two in our case.

The classification of the LCS/LCM uses, or application areas, proposed by the EPA features five tiers (Tier I–Tier V), presented in Table 1, which depend on the pollutant considered, the MNB and CV values.

Once the CV and MNB have been computed for a sensor model, it is possible to determine the optimal "tier", or application area, of the LCM/LCS. The conditions necessary to assign an LCS/LCM model to a tier are determined by the "AND" logic operation for the ranges of the MNB and CV values indicated in Table 1. Thus, as an example, an LCM/LCS model belonging to Tier IV must demonstrate both $-0.3 <$ MNB < 0.3 and CV < 0.3.

Table 1. Classification of possible uses, or application areas, for LCSs/LCMs proposed in EPA guidelines.

Tier	Application Area	Pollutants	MNB	CV	Application Examples
I	Education and information	All	$-0.5 < MNB < 0.5$	$CV < 0.5$	Providing informal information about the presence of a pollutant; the use of sensors as teaching tools
II	Hotspot identification and characterization	All	$-0.3 < MNB < 0.3$	$CV < 0.3$	The identification of emission sources of pollutants such as heavy traffic or industrial facilities
III	Supplemental monitoring	O_3, NO_2, PM, CO, SO_2, and TVOCs	$-0.2 < MNB < 0.2$	$CV < 0.2$	Supplementing the regulatory network monitoring for improving the spatio-temporal resolution of pollutant maps
IV	Personal exposure monitoring	All	$-0.3 < MNB < 0.3$	$CV < 0.3$	These sensors can be used in mobile monitors of a size that can be easily carried by users for measuring pollutant concentrations in indoor/outdoor environments
V	Regulatory monitoring	O_3 CO, SO_2, PM_{10}, $PM_{2.5}$, and NO_2	$-0.07 < MNB < 0.07$; $-0.1 < MNB < 0.1$; $-0.15 < MNB < 0.15$	$CV < 0.07$ $CV < 0.2$ $CV < 0.15$	Pollutant monitoring to determine if an area complies with the national ambient air quality standards

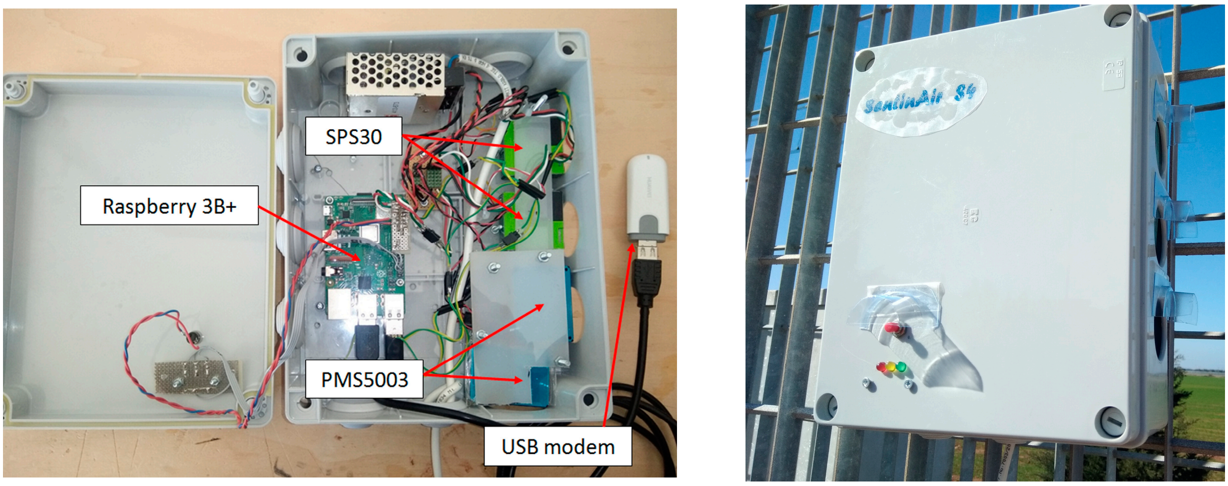
4. Materials and Methods

The LCSs available on the market for air quality monitoring are devices able to measure pollutant concentrations using different technologies and working principles. Nonetheless, aside from the sensor type and technology used, these devices need appropriate electronic boards for their effective use. The electronic circuitry of these boards converts the current or the voltage output of the sensing element, dependent on the pollutant concentration, into an electronic signal available at the output interface. These sensors can present various types of interfaces, the most common being: analog, TTL serial, I2C, and USB. To effectively make use of the data produced by the sensors, it is required a suitable electronic system capable of reading the signals coming out of the output interfaces and converting them into usable data. The overall electronic system in charge of accomplishing this task is commonly called a low-cost monitor (LCM).

The PM sensors evaluated in this work were integrated into the SentinAir platform [34] designed to act as both a tool for quickly performing evaluations/calibrations of LCSs and an LCM for indoor or outdoor environments [21,35–37]. The SentinAir system is an in-house and open-source design implemented in the ENEA research center of Brindisi, located in the Puglia region of Italy. Therefore, all the materials, software, and procedures required to build a copy of SentinAir are available online on the project repository webpage [34].

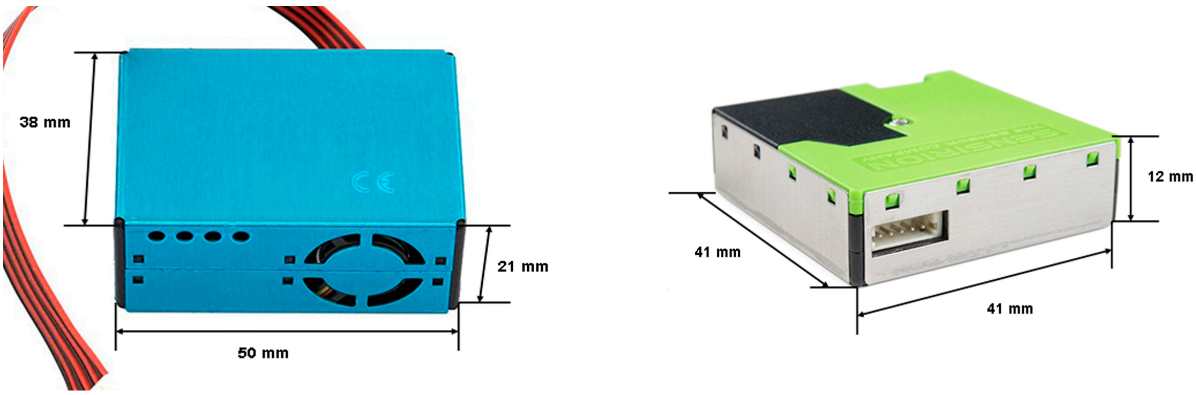
The main difference between the SentinAir system and the other commercially available LCMs is represented by the possibility of integrating, or installing, a huge variety of sensors presenting any of the earlier mentioned output interfaces and produced by various manufacturers. This capability was achieved thanks to the adoption of the low-cost micro-computer Raspberry 3B+ [38] as the core of the system and the software created for the operation of SentinAir. As a matter of fact, the Raspberry 3B+ hardware is characterized by four USB ports, an I2C bus, and a TTL serial port, while a driver software for the installation of an analog-to-digital converter (ADC) was designed and implemented to allow the use of the ADCPi board by ABelectronics [39]. This electronic board is necessary for the use of the LCSs with an analog output interface. A software system composed of drivers written in the Python language [40] specific to each device or sensor installable in SentinAir provides the "plug-and-play" feature for the system. Another very useful feature of SentinAir consists in its ability to be remotely operated through its dual wireless communication system: a WiFi channel and an internet connection via a USB modem.

The SentinAir device used for the test acted as an LCM containing two copies of the PMS5003 sensor model and two copies of the SPS30 model, as illustrated in Figure 1. The LCSs used in this experiment were particle optical counters available on the market, whose hardware was contained in a compact case, as shown in Figure 2.



(**a**) (**b**)

Figure 1. The SentinAir platform: (**a**) the device used in the experiment with the four PM sensors installed inside; (**b**) external view of the device acting as an LCM in this experiment.



(**a**) (**b**)

Figure 2. The miniaturized sensors used in this study: (**a**) the PMS5003 produced by Plantower and its size; (**b**) the SPS30 produced by Sensirion and its size.

Their working principle is illustrated in Figure 3. It is based on a laser beam scattered by the particles entering a detection camera. The more particles cross the laser beam, the more scattered the beam becomes. The beam is detected by a light detector that provides an electronic signal depending on the scattering level of the laser light. Thus, the microprocessor inside the sensor is devoted to translating the electronic signal of the detector into numbers of particles per volume unit. Therefore, an algorithm implemented in the sensor microprocessor provides the PM concentration depending on the number of particles detected. Unfortunately, the manufacturers do not provide details about this algorithm. Concerning the output interface, these LCSs are provided with a serial TTL interface in the case of the PMS5003 sensor, and an I2C or a serial TTL interface in the case of the SPS30 sensor. By connecting the LCM hardware through these interfaces, measurements related to PM_{10}, $PM_{2.5}$, and PM_1 (and, in the case of SPS30, also PM_4) concentrations can be read. More details about these LCSs can be found by downloading the datasheets from the manufacturer's websites [29,30].

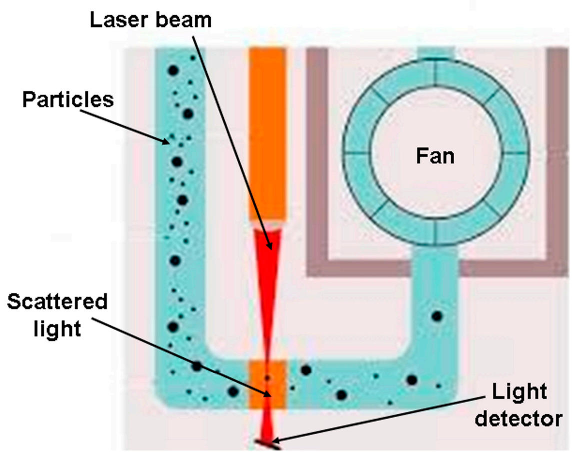


Figure 3. The simplified scheme of an optical counter for PM detection on which the working principle of LCSs used in this research is based.

Other LCMs similar to SentinAir are currently available on the market. These are all based on LCSs whose working principle follows the simplified scheme depicted in Figure 3. An estimation of their costs is provided in Table 2, where some examples of both LCMs and LCSs for PM measurements are listed.

Table 2. A summary of some LCMs/LCSs for PM measurements currently available on the market with their indicative costs and manufacturers.

Device Name	Cost (EUR)	Manufacturer	Device Type
PMS5003	~20	Plantower	LCS
SPS30	~50	Sensirion	LCS
OPC-N2	~350	Alphasense	LCS
SDS011	~33	Nova Fitness	LCS
PurpleAir PA-II	~190	PurpleAir	LCM
Airly PM	~900	Airly	LCM
Airquality Egg 2022	~630	Airquality Egg	LCM
TSI Bluesky	~380	TSI	LCM

The experiment designed to assess the potential uses of these sensors consisted in placing a SentinAir device near the ARPA Puglia fixed monitoring station located in the town of Mesagne (Italy), as shown in Figure 4.

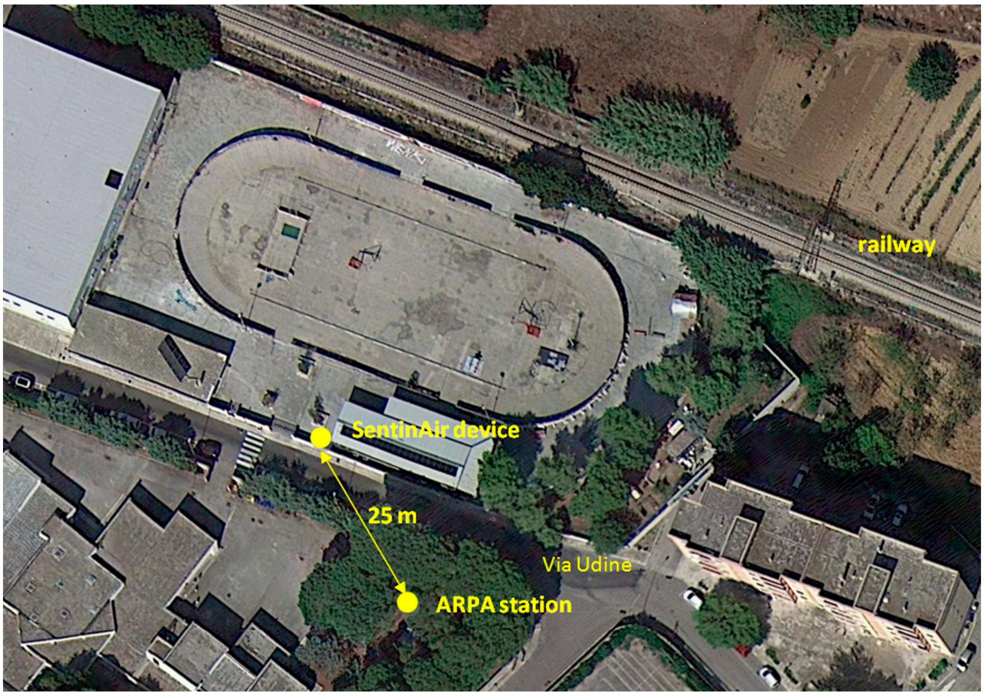


Figure 4. Depiction of the site where the ARPA station was located.

The reference used for the evaluation was the ARPA Puglia monitoring station located in Via Udine, which is devoted to measuring the background concentrations of PM_{10} in a suburban environment. It performs the measurements using gravimetric instruments and provides data concerning the daily means of this pollutant that are freely downloadable from the ARPA website [32].

Data elaboration and indicator computations were performed using the Scikit-learn libraries written in the Python language [41–43], which are an open-source software freely downloadable from their website [41]. The sensor data were read by the SentinAir device every 5 min, and, subsequently, daily means were computed for building the database together with the ARPA data. Italian rules for PM monitoring require just the monitoring of PM_{10} on a daily average basis and yearly averages for $PM_{2.5}$; for this reason, the useful data provided by the fixed reference station were only those related to the daily averages of PM_{10}. The SentinAir system can store the data related to the performed measurements on its internal SD card memory; moreover, it features a web server from which the user can download any data measured by the device thanks to the presence of a USB modem [35–37]. The data gathered by the SentinAir device were periodically downloaded and joined with the ARPA measurements that are publicly available on the ARPA website [32]. Starting from the daily averages of the ARPA monitoring station and the daily averages computed by the LCM, it was possible to assess the LCS uses related to PM_{10} monitoring by considering the MNB and CV values computed for each sensor model according to the EPA guidelines [13,19].

The performance of PM sensors is negatively affected by the environmental relative humidity [44,45]. The reason for this effect is the condensation of water vapor that makes the aerosol particles comprising the particulate matter grow hygroscopically. This phenomenon causes incorrect measurements in devices based on the optical counter working principle, such as the PM sensors considered in this experiment. Some studies [44,45] have proposed an algorithm to take into account the negative effect of humidity in order to improve sensor performance. This algorithm consists in applying a correction factor to the sensor readings, as indicated by the following formulas:

$$PM_{corrected} = \frac{PM_{uncorrected}}{C} \tag{6}$$

$$C = 1 + \frac{\frac{k}{1.65}}{\frac{100}{RH} - 1} \tag{7}$$

where RH is the relative humidity, and "k" is a parameter depending on the nature of the particulate matter. To have a complete view of the capabilities of the sensors considered in our investigation, we applied the correction algorithm illustrated above to understand the extent to which the performance of the sensors could be improved. However, in our experiment, no information could be collected on the composition of PM compounds; therefore, we selected two distinct values (k = 0.5 and k = 0.62), as suggested in the works of Crilley [44] and Di Antonio [45]. In the first study, it was stated that the expected range of the "k" parameter could be reasonably thought to be 0.48–0.51 for PM_{10}, while the second study hypothesized that the "k" value could be set equal to 0.62 in the case of a mixture of organic and inorganic compounds, such as in a typical polluted urban environment.

5. Results

For the purpose of this study, measurements from the 15th of September 2022 to the 27th of November 2022 were carried out. These data formed a dataset composed of the daily averages of PM_{10} concentrations resulting from the measurements of the four LCSs involved in the study (hereafter named "PMS5003(1)", "PMS5003(2)", "SPS30(1)", and "SPS30(2)") and the fixed reference station of ARPA Puglia (hereafter named "reference"). In the period indicated above, ARPA Puglia did not provide data for ten dates due to the maintenance of the instruments or other unknown issues, while the data recovery percentage was close to 100% for each sensor. For this reason, we excluded the records of the database lacking reference measurements from the computation of the performance indicators.

5.1. Results of the Performance without Considering the Humidity Effect

One of the aims of this study was the characterization of "out-of-box" PM data offered by the four sensors available on the market; thus, in this first stage, we assumed that the factory calibrations performed by the manufacturers would reflect the PM concentrations in the best way.

In Figure 5, the time series related to the LCSs under evaluation are reported along with the measurements of the reference. In this figure, it can be noted that higher PM_{10} concentrations were more frequent in the latter period of the experiment. We could explain this element by considering that, as the colder days approached, the use of wood burners, which are widely employed in the town for domestic heating, became more frequent. By the examination of this figure, it can also be noted that, in general, the SPS30 model tended to underestimate the PM_{10} concentrations, while the PSM5003 model tended to overestimate them. Nevertheless, we observed substantial agreement between the measurements of the two copies of the sensors for each of the two models. Another finding shown in Figure 5 was that the PM_{10} concentrations measured by the reference ranged from 7 $\mu g/m^3$ to 51 $\mu g/m^3$. Concerning the on-field evaluation of LCSs, wider ranges of pollutant concentrations increase the probability of reporting a better LCS performance in terms of R^2. In this regard, the maximum value registered during the experiment was barely higher than the limit fixed by the Italian regulations, which is 50 $\mu g/m^3$. This value was reported on only one day, the 26th of November 2022. The average PM_{10} concentration measured by the reference was equal to 20.4 $\mu g/m^3$, less than half the limit fixed by the Italian regulations.