

Article

Not peer-reviewed version

Future Low-Cost Urban Air Quality Monitoring Networks: Lessons from the EU-UIA AirHeritage project

[Saverio De Vito](#)*, [Antonio Del Giudice](#), [Gerardo D'Elia](#), [Elena Esposito](#), [Grazia Fattoruso](#)*, [Segio Ferlito](#), [Fabrizio Formisano](#), Giuseppe Loffredo, [Ettore Massera](#), [Paolo D' Auria](#), [Girolamo Di Francia](#)

Posted Date: 6 August 2024

doi: 10.20944/preprints202408.0389.v1

Keywords: Pervasive Air Quality Monitoring, Citizen Science, Air Quality Research Projects, Pervasive Air Quality Monitoring Dataset, Low Cost Air Quality Multisensors Systems, Artificial Intelligence, Calibration



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Future Low-Cost Urban Air Quality Monitoring Networks: Lessons from the EU-UIA AirHeritage Project

Saverio De Vito ^{1,*}, Antonio Del Giudice ¹, Gerardo D'Elia ¹, Elena Esposito ¹, Grazia Fattoruso ^{1,*}, Sergio Ferlito ¹, Fabrizio Formisano ¹, Giuseppe Loffredo ¹, Ettore Massera ², Paolo D' Auria ³ and Girolamo Di Francia ¹

¹ TERIN-SSI-EDS Laboratory, ENEA CR-Portici, P. le E. Fermi 1, 80055 Portici, Italy

² TERIN-SPV-DIN Laboratory, ENEA CR-Portici, P. le E. Fermi 1, 80055 Portici, Italy

³ Agenzia Regionale per la Protezione dell' Ambiente – Campania, Naples, Italy

* Correspondence: ; saverio.devito@enea.it (S.D.V.); grazia.fattoruso@enea.it (G.F.)

Abstract: The last decade has seen a significant growth in low-cost air quality monitoring systems (LCAQMS) adoption, mostly driven by the spatial density limitations of traditional regulatory grade networks which restrict their effectiveness in several settings. The EU's regulatory framework advise for deploying one air quality monitoring unit per roughly 170,000 people or 15 square kilometers which is insufficient for capturing high spatio-temporal variance in complex urban scenarios. Often these settings are characterized by distributed emissions in small roads subjected to intense cars traffic, house heating driven emissions and frequently subjected to canyon effects. The Air Heritage project in Portici, funded by the EU's Urban Innovative Actions (UIA) program, addressed these challenges by integrating low-cost sensors with traditional monitoring systems and engaging the community in pervasive measurement strategies. The project highlights the importance of individualized sensor calibration and frequent recalibrations to maintain accuracy. It also showcases the benefits of citizen science approaches, IoT infrastructures, AI components, and geostatistical sensor fusion algorithms for mobile and opportunistic air quality measurements. This paper reports on the lessons learned and experiences from the Air Heritage project, aiming to guide future implementations in similar contexts. By sharing these insights along with the gathered datalake, we aim to inform stakeholders, including researchers, public authorities, and citizens, about effective strategies for deploying and utilizing LCAQMS to enhance air quality monitoring and public awareness.

Keywords: pervasive air quality monitoring; citizen science; air quality research projects; Pervasive air quality monitoring dataset; low cost air quality multisensors systems; artificial intelligence; calibration

1. Introduction

Last decades have seen the raise of awareness and concerns regarding the Air quality problem [1]. Air quality concerns are particularly perceived in densely populated urban areas where they can significantly and largely impact public health by exposing them to harmful pollutants [2,3]. It is well known, actually, that their inhalation can cause several affections impacting respiratory organs (e.g. COPD, [4]) , circulation (e.g. stroke) and cause development of cancer [5]. In particular, NO₂ short term exposure is known to be a cause of severe asthma while long term exposure is usually associated with several respiratory diseases [6]. Ozone is instead well studied as irritant of the upper respiratory ways for short term exposures while there are evidences of long term effects in terms of accelerated decline in lung function and the progression of emphysema [7]. Particulate matter depending on their size can be found in several tissues where it accumulates and can be source of continuous oxidative stress leading to ischemia [8] and cancer [9]. Hence, the increasing the pervasivity of accurate, factual knowledge about AQ concentration is a high valued goal. It can give political decision makers and

citizens the ability to act based on scientifically validated data with positive impacts on quality of life, public health further reducing environmental inequalities.

One of the first driver of environmental inequalities is, actually, the scarcity of air quality information in localized areas of cities [10]. Pervasiveness, however, is out of the operational envelope of the accurate but costly and bulky regulatory grade AQ analyzers causing a number of research resources to focus on low cost air quality monitor device development. During the last years, a significant number of funded research projects have been devoted to the development of technological and methodological strategies for Air quality pervasive and accurate monitoring (see OpenSense; Citi-Sense-MOB; Everyaware; Citi-Sense; "Schools for Better Air Quality" UNICEF Serbia pilot initiative, Breath London Network, [11–15]). These complex projects face a number of difficult design choices trying to accommodate technological and social sciences opportunities and limits with their general requirements and goals. They could target pervasive monitoring, citizen engagement, AQ information augmentation, traffic or heating regulation, etc. or any combination of these. Of course these will inform and lead the solution choice. When accuracy is the main concern, for example, pinpointing the right sensing technology base is fundamental though it may conflict with financial resources [16]. Data processing pipeline is fundamental to achieve the sought accuracy and calibration methodology choice is certainly at the core. Field calibration is the state of the art of LCAQMS calibration methodology but this choice can generate logistic issues when the deployment node is foreseen to exceed several tenths. UNICEF Serbia have recently started a pilot project with more than 120 deployed units [14]. Acceptability and easy of use of sensor nodes will be paramount when citizen science is targeted [17]. Eventually, preparing the citizen to correctly grasp the meaning of measurements is a difficult task needing adequate resources as well as communicating AQ data through universally understandable, accuracy informed and concise index without generating false expectations [18,19]. This of course depends on the cultural substrate, legal framework and expectation. If legal framework accounting is based on number of concentration thresholds overshoots than it could be easy to resort to this indicator. However, a common expectation is to use LCAQMS nodes to drive the local administration to act for remediation. Of course this is a false expectation since only regulatory grade monitoring device measurements are legally binding.

Most of the projects focused on fixed AQ network development but the development of portable AQ monitoring device is equally important for obtaining personal exposure insights. Low cost is a very important feature when aiming to personal monitors and is surely to be taken into account when targeting sustainable strategies for low income countries where highest concentrations for some pollutant can be recorded. Conversely, accuracy is the most sought feature when increase of regulatory grade network spatial coverage is the main goal.

A growing trend can also be observed for the number of devices currently involved in those projects aiming to involve a large number of the citizens. They could be either citizen scientists cooperating to the monitoring activities or observers participating in reviewing the results in seek for enhanced awareness and behavioural changes. In this framework our group observed the peculiarity of the Italian and European scenario while selecting common worldwide requirements. Usually, our urban settlements are mostly impacted by pollution generated by car traffic and household heating. The high density of historical features in city centers characterized by small spaces, the European legal framework and costs consideration prevented the development of pervasive air quality networks. As such we recognized the role of citizens awareness while keeping the momentum towards accurate measurements. For this reason we involved our partners in the development of the AirHeritage project. This has been a Urban Innovative Action project funded by EU under ERDF framework to be implemented in Portici, a small high population density city located 7km south of Naples in south Italy [20]. Main aim was increasing spatial and temporal density information about most common air pollutants and simultaneously involving citizens associations in measurement process by deploying 30+ mobile devices thus increasing their knowledge on personal exposure, further densify the measurement process and increase their awareness. Aside from the measurement approach, a high resolution AQ forecasting model based on fluidynamic simulation, weather forecast data and measurement data assimilation developed by Terraria srl and Prof. Mazzoleni group from

University of Naples have been deployed. Several informative and preparatory campaigns with citizens and schools have been conducted with the support of Legambiente Campania local associations and the factual involvement of local administration authorities. This innovative, integrated, cooperative approach has strongly characterized the project ideation and development and ultimately guaranteed its technical and social success, while we have to recognize that a thoroughly effective pollution mitigation policy needs a continued and sustained in time approach. This work actually focus on sharing the lesson learnt on the technological side of the design and operation of a pervasive and accurate measurement approach culminating in the development of a hybrid network integrating fixed/mobile/regulatory grade analyzers. Next chapters focus respectively on the development of the AirHeritage version of the ENEA MONICA™ device, their lab based characterization and the development of a full-blown IT infrastructure for AH network data gathering, processing, storage and visualization. Furthermore, an efficient colocation strategy at the base of the adopted field calibration and validation approach is described. Eventually, the use of the gathered opportunistic measurements to generate ultra-high spatial resolution concentration maps is described completing the technological and sensors data processing pipeline. Each of the steps are completed by shorts and focused sentences highlighting the lessons learnt during the project implementation phase.

2. The MONICA Low-Cost Air Quality Device

The MONICA device design (shown in Figures 1 and 2) lies at the core of the Air Heritage Hierarchical network. It showcases a successful approach to mobile air quality monitoring with a focus on extended operational life and efficient data collection in highly pervasive monitoring scenarios. This section dives deeper into the design choices and technical components that enabled this achievement. The functional requirements described a system that could be easily customized for both fixed and mobile monitoring scenario. Weight and form factor where of course paramount in allowing citizen to easily carry the unit while at the same time not sacrificing accuracy. Accordingly, three main aspects guided the technical design choices: Sensor selection, optimization of internal operations and power efficiency, data transmission.

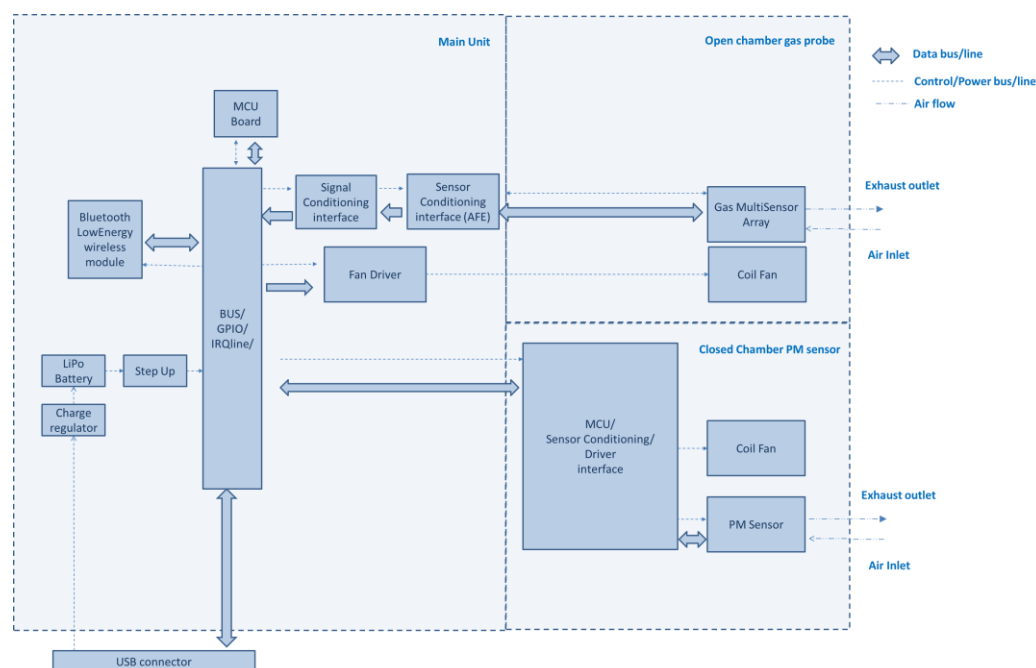


Figure 1. MONICA(TM) node diagram.

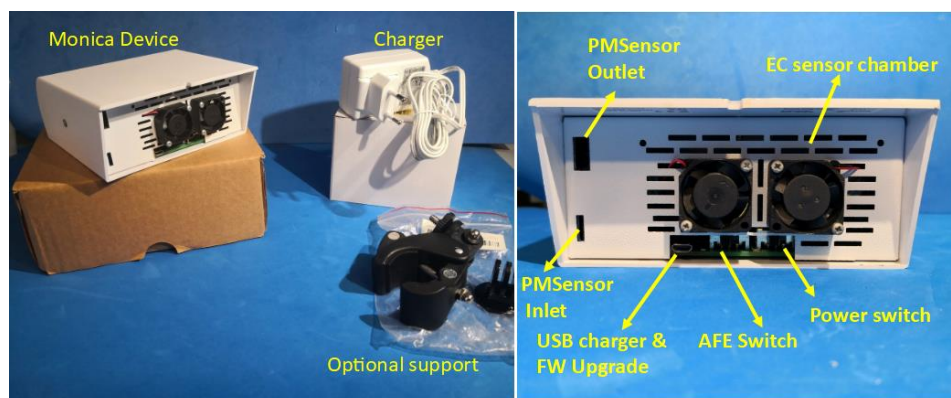


Figure 2. Front and back picture of the MONICA node.

2.1. Sensor Selection

The MONICA device is mostly based on Alphasense™ electrochemical sensors [21]. These sensors were selected for their low-power operation while maintaining accurate measurement of core air pollutants, including carbon monoxide (CO; sensor CO-A4), nitrogen dioxide (NO₂; sensor NO2-A43F), and ozone (O₃; sensor OX-A431). This enables valuable insights into potential health risks associated with these pollutants to be gained. The raw current signals generated by the electrochemical sensors are processed by an Alphasense™ AFE 810-0020-00. This component converts these signals into voltages for easier acquisition by the data processing unit. Additionally, a dedicated temperature and humidity sensor (Silicon Labs SI7022) is located in close proximity to the electrochemical sensors. This data is of paramount importance for the correction of potential measurement errors resulting from environmental fluctuations. The Plantower PMS7003 sensor complete the sensor suite, enabling the MONICA device to measure particulate matter (PM) fractions of PM_{2.5} and PM₁₀ [22]. The sensor boasts an operating range of 0-500µg/m³ with a resolution of 1µg/m³, providing detailed insights into the concentration of harmful airborne particles. It is noteworthy that the sensor incorporates an autonomous fan, which ensures accurate airflow and, consequently, optimal PM detection.

2.2. Internal Operations and Power Efficiency

Data acquisition is handled by a low-power 32-bit STM32(STM32L432KC) microcontroller from STMicroelectronics™. This choice balances processing power with minimal power consumption. The STM32 Nucleo™ LK432KC board with a 12-bit ADC takes the processed voltage signals from the AFE and translates them into digital data for further analysis. The ADC resolution (and outputs AFE range matching) allows for the acquisition of samples representative of concentrations compatible with that required by typical air monitoring devices (0–500 ppb). In the special case of CO, it is possible to sense values in the entire sensor range extension (0-12ppm).

To further optimize power consumption, the MONICA device features intelligent sleep states for various components. This ensures that they only consume power when actively involved in sensing or data processing tasks.

The impact of fan operation on power consumption has also been taken into account. In fact, while airflow is critical to sensor function, fan selection and operation have a significant impact on power consumption. In fact, part of the metal casing of the instrument acts as a Faraday cage to shield the EC sensors from EMI. This is critical for gases such as NO₂ and O₃, which have poor air mixing and instability in partially open chambers. Fans address this by promoting gas-air mixing and mitigating instability through continuous air injection. The MONICA instrument uses a dual approach: fixed speed fans (ON/OFF switchable) for EC sensors with minimal airflow requirements to limit this problem, and autonomous fans for sensors requiring precise airflow control (PM sensor) to ensure optimum power consumption without compromising data accuracy.

A 3800mAh Li-Po battery and a light and compact design provide the basis for extended mobility operation. This means at least 8 hours of continuous air quality monitoring without recharging, a significant advantage for mobile applications.

2.3. Communication and Data Transmission

The sensor node uses BLE technology (BLE module HM-10) to communicate with modern smartphones transmitting JSON payloads. This efficient wireless connection serves multiple purposes:

- Geo-referencing: The smartphone's GPS data can be used to geotag the collected air quality measurements, providing critical spatial context for the collected data.
- Calibration implementation: Calibration algorithms can be run on the smartphone to ensure the accuracy of sensor readings over time.
- Data transmission to the IoT backend: The smartphone acts as a bridge, seamlessly transmitting the collected and processed air quality data to the central IoT backend for further analysis and visualization.

The architectures and details of user-side operation of sensors data transmission are explained in the following sections.

2.4. Benefits of the MONICA design

The design of the MONICA device demonstrates the feasibility and effectiveness of creating a mobile air quality monitoring system that prioritizes low-power operation and efficient data collection. This approach offers several key benefits:

- Extended Mobile Operation: The low-power and comfortable to wear design choices, ensure the MONICA device can function for extended periods without needing a recharge, making it ideal for mobile deployments in diverse environments.
- Accurate and Comprehensive Monitoring: The combination of various sensors technologies allows for the measurement of a broad range of air pollutants, providing a more holistic picture of air quality.
- Seamless Data Collection and Transmission: BLE connectivity with smartphones facilitates convenient data collection, georeferencing, calibration, and transmission to the central data storage and analysis platform.

By implementing these strategies, the MONICA device paves the way for more widespread and efficient mobile air quality monitoring, empowering researchers and policymakers with valuable data to address air pollution challenges in our cities.

3. Laboratory Characterization of MONICA Devices

3.1. Procedures for Laboratory Characterization and Calibration

Eleven Monica devices were tested in a laboratory under controlled atmosphere conditions to verify their proper functioning and measure the performance of the sensors by comparing them with the specifications stated in the manufacturer's calibration documents. The accuracy of the laboratory tests is ensured by the high quality of the instrumentation used and the laboratory's twenty years of experience in studying solid-state gas sensors in controlled atmospheres. Specifically, for validating nitrogen dioxide concentration measurements, which is a very aggressive and difficult-to-control gas in enclosed environments, we used EPA-certified instrumentation (Teledyne T200 [23]). More than 50 tests were conducted, implementing measurement protocols to optimize the accuracy and precision of the tests and minimize measurement times. The equipment and measurement protocols are briefly described below.

A 15-liter "Large Volume Test Chamber" (LVTC) has been installed in a state-of-the-art Gas Sensor Characterization System (GSCS). The GSCS consists of a stainless steel, air-tight test chamber enclosed in an adjustable thermal box. Within the LVTC, the air composition (humidity, chemical

compounds concentration) can be precisely controlled through an inlet of gas flux regulated by certified Mass Flow Controllers (MKS 1179 series). The accuracy of the gas's chemical composition is ensured by mixing gases from accredited bottles (Rivoira / Nippon Gases). Temperature and humidity are monitored with industrial sensors (LSI Pt100). The LVTC is capable of calibrating multiple sensor systems simultaneously.

The calibration method involves injecting a constant flow of the target gas, properly diluted to the maximum concentration (C_0) with humid synthetic air, into the inlet tube of the LVTC. The time-dependent concentration $C(t)$ of the target gas is accurately predicted by the following exponential law, which generally describes the transition between two steady states of a physical parameter under a time-constant perturbation:

$$C(t) = C_0 \left(1 - e^{-\frac{t}{\tau}}\right) \quad (1)$$

A calibrated sensor can precisely estimate the characteristic time (τ). This parameter is proportional to the dead volume inside the chamber and must be appropriately corrected when multiple sensors are inside the chamber.

The calibration procedure consists of three steps: first, synthetic air is injected to establish the unperturbed state, recording the sensor output as a baseline; in the second step, the target gas, properly diluted in the gas carrier, is injected, and the adsorbing phase of the sensor response is recorded; finally, the third step involves washing the test chamber with a constant flow of synthetic air, recording the desorbing phase of the sensor output. This procedure allows verification of the sensor output behavior during the adsorbing and desorbing phases of the chemical compounds on the sensor surface. Sensing hysteresis or poisoning can be detected and measured.

With a gas flow rate of 1 liter per minute, the τ of the LVTC is estimated to be 1100 seconds \pm 50 seconds. This means that in 3 hours, several sensors with a response time faster than 2 minutes can be calibrated with maximum precision from 0 to C_0 .

For each tested device, the sensitivity to CO and NO₂ is verified, while for Ozone sensor only the sensitivity versus NO₂ was tested.

3.1. Results and Performance Improvement Suggestions

One of the most important aspect to control is the impact of sensor node case mounting on its performances. Worsened air mixing may impact on sensitivity and responsiveness. Furthermore, vendor reported sensitivity and LOD characteristics should be checked to ensure the requirements fulfillment when mounted. For these reasons We strongly advise to check the node performance in laboratory controlled settings in order to capture the main issues that may affect the sensor when field deployed. Checking interferences impact is also very important to decide about the calibration strategy and its parameters. Actually, the EC Sensors installed on Monica meet the expected performance standards and highlighted the very well known temperature dependence.. More importantly, laboratory tests highlighted the need to equip the devices with a fan in front of the NO₂ and NO₃+O₃ sensors, having observed a loss of efficiency in measuring NO₂ due to the case. The fan speed was set to maximize the sensitivity of the sensors while minimizing power consumption.

4. Comprehensive Software Platform Development

The Air-Heritage project had to handle different air quality data sources and provide many functionalities to the citizens, which required a careful design for the software platform to fulfill all the requirements. The air quality sensors, selected for deployment during AirHeritage project were of two different types, with different sampling and data transmission methodologies:

1. ENEA MONICA sensors nodes operated at high sampling rates (about 3.5 seconds min sampling period set to 6s)
2. Commercial Fixed stations by Digiteco srl [24] with data sampled every minute and transmitted every 15 minutes

As above mentioned, the MONICA™ sensors were fully developed at ENEA Portici Research center. These devices are not equipped with wide area network connectivity and has no autonomous

time clock reference. Consequently, as previously stated, data acquired from the devices must be sent via BLE to a device capable of transmitting the data to a remote server for centralized storage. The data must be augmented with a correct timestamp (in UTC format and synchronized with a time server). This device was either an Android smartphone using a dedicated App (for the mobility campaigns) or a custom sw equipped Raspberry Pi connected to wifi/cellular router.

Data generated from MONICA™ is actually sent using BLE to the WAN-enabled device in JSON format. Only raw, uncalibrated data is transmitted. This same format is used to send data using the HTTP protocol to the remote server. Raw data from MONICA devices is augmented with calibrated data. Only PM concentration values are available directly from the MONICA devices as expressed in $\mu\text{g}/\text{m}^3$ (using Plantower sensors vendor calibration), and the calibration is done on the edge device (i.e., an Android smartphone, although the same is possible on the Raspberry Pi).

Data from the fixed stations used in the project, are sent via GPRS connectivity to an MQTT broker managed by the device supplier. The data is collected every minute and transmitted using the MQTT protocol every 15 minutes.

Data from the fixed stations comprise raw data as well as calibrated data. The initial calibration was provided by the device supplier using the default calibration from the sensor manufacturer. Later, after a co-location calibration, ENEA developed and compared multiple calibration methodologies, both statistical, such as Multiple Linear Regression (MLR) and AI-based (such as Random Forest, RF), significantly improving the provided measurements accuracy, especially using RF [25–28].

Almost all data involved in the project, with different time sampling, is essentially always time series data that could, theoretically, generate a huge amount of data and traffic towards the remote server responsible for storing all data in a centralized way.

Other requirements for the project were:

- Provide a user interface, preferable web based, to allow clients to:
 - Store and retrieve MONICA sessions, but only for registered users
 - Report, in an interactive map, the status of the fixed stations with the AQI that should be updated, according to reglementary specifications, every 15 minutes (the time interval for an update from the MQTT broker)
 - Download data as CSV format from fixed station or for specific session
- Store data in a database in an effective way being able to keep good performances as time should passed by even in case of multiple devices and for long lasting MONICA sessions
- Fulfill FAIR (Findability, Accessibility, Interoperability, and Reuse) requirements to allow easy access to data [29].
- Consider best practices in exposing services.
- Address concerns about cybersecurity, including authentication and authorization.

4.1. Architecture of the Developed Software Platform

Taking into consideration all the above, after a deep analysis of the requirements and best practices involved in each field, the design phase resulted in the architecture depicted in Figure 3 below.

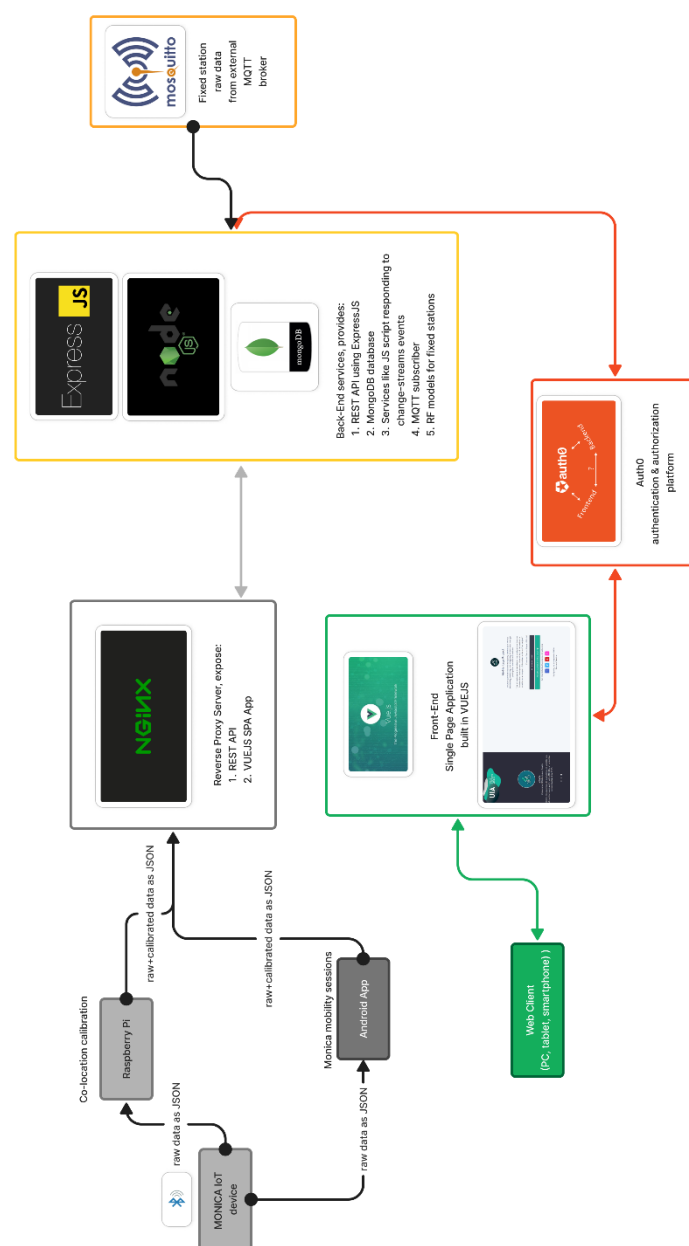


Figure 3. Synthetic schema complete software architecture Air Heritage Project.

As clearly reported in Figure 3, air quality data is stored on a remote server using a REST (REpresentational State Transfer) API service [30], through CRUD endpoints (Create, Read, Update, Delete) exposed by NGINX, a well-known reverse proxy server [31]. Data is sent from the Android app/Raspberry Pi using the HTTP protocol in JSON format [32]. The REST API service was developed using an opinionated web framework, ExpressJS, based on NodeJS [33]. Some REST API endpoints, mainly that related to CRUD operations, were protected using OAuth2 authorization Protocol [34]. This means that the generic client willing to use the described endpoints had to sign up for a valid account before being able to access these protected endpoints. For this task, the authors decided to use an external authentication and authorization service, Auth0, which offers both free and paid tiers [35]. This approach made it possible to implement the following authentication mechanism:

- Social login as Facebook, Google, Twitter (now X)
- Account defined on Auth0 servers

This allowed to implement authentication and authorization in a simple and rapid way and support M2M (Machine to Machine) authentication for the Raspberry Pi devices [36].

Data from fixed stations were inserted into the database by a NodeJS application acting as an MQTT subscriber. Every 15 minutes, a message with proprietary binary format was sent by the MQTT broker. Once correctly parsed, it was saved into a dedicated collection in MongoDB.

Initially, data from fixed stations included both raw and calibrated values for pollutants such as CO, NO₂, and PM₁₀, although the calibration reflected only the sensor manufacturer's indications. After an appropriate co-location procedure, calibration models developed by ENEA, particularly a Random Forest (RF) model, were deployed. The new calibrated values were computed using ad hoc JavaScript code executed on the server via NodeJS. By utilizing MongoDB's change-stream functionality a query automatically extracted inputs for the RF model and updated the collection with the newly calibrated values whenever new values were inserted into the respective collection [37]. A similar procedure was used, via another JavaScript script, to compute the Air Quality Index (AQI) and update it every 15 minutes. The services executed on servers hosted at the ENEA Research Center constitute what are commonly known as backend services. The user interface, also known as the frontend, was developed using a JavaScript framework with associated functionalities (Vuetify as the UI library, VUE I18N for internationalization, etc.). The decision to develop the front-end application using Vue.js (version 2.x during the project's duration) was made due to its progressive performance and versatility in building web user interfaces. Other common alternatives include ReactJs or AngularJS. This choice enabled the development of a Single Page Application (SPA) for the project [38].

Figure 4 shows the home page of the site available at <https://airheritage.portici.enea.it>. Note the choice to proceed as a guest or a registered user; the latter option allows you to sign in or create an account using Auth0 services configured for the Air-Heritage Project. Authentication and authorization were handled by Auth0, and the user interface for signing in or creating an account was the same on both Android and web browsers (see Figure 5).

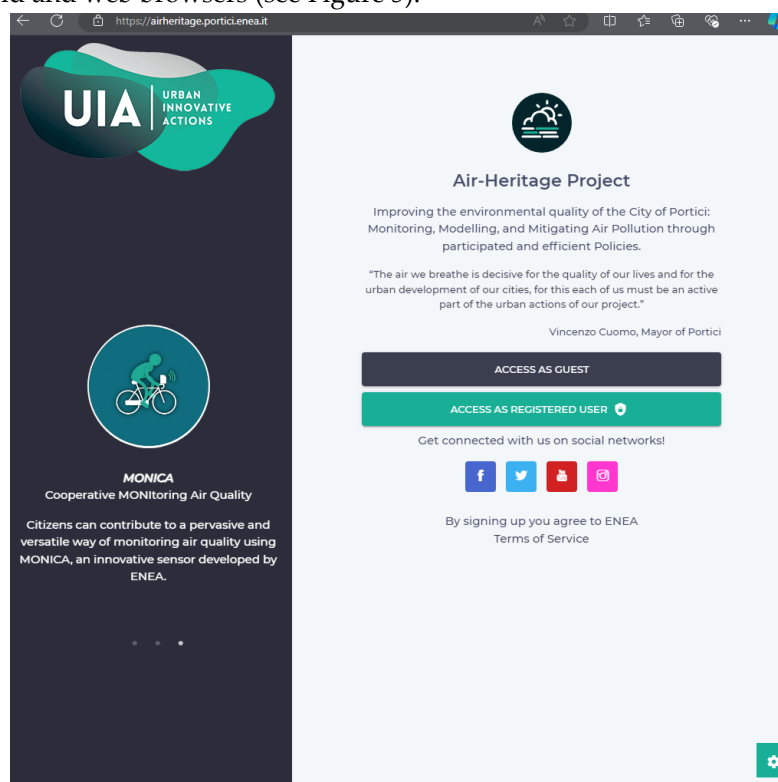


Figure 4. Air Heritage home page.

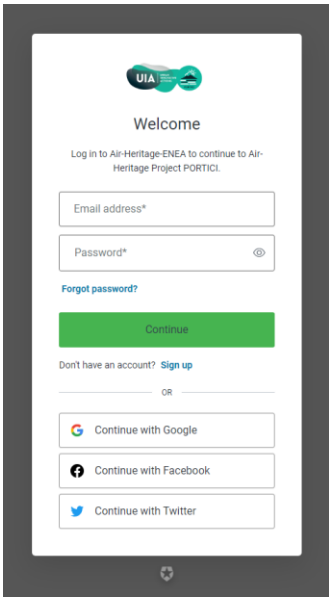


Figure 5. Sign in/up provided using Auth0 service.

The front-end reported general information about the project and allowed users to view logged users' registered MONICA sessions. It also displayed real-time data from fixed stations in terms of AQI (Air Quality Index) on an interactive map (created using LeafletJS), with updates every 15 minutes [39,40].

Figures 6 and 7 provide examples of the UI built for the project related to the two main features described earlier.

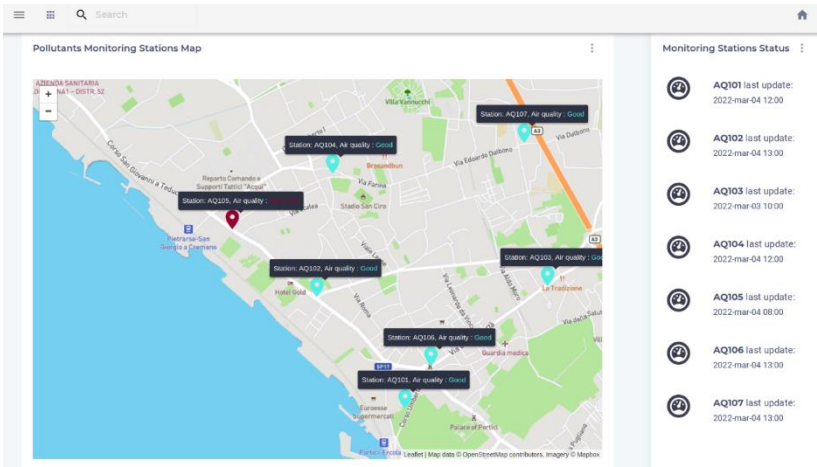


Figure 6. Status of Air quality from fixed stations.

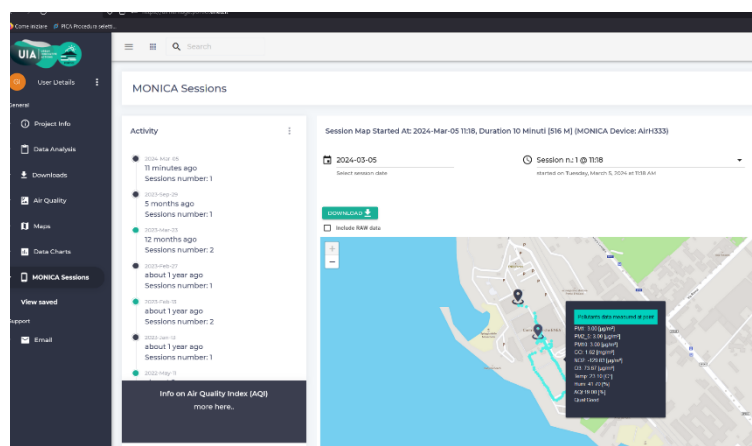


Figure 7. Interactiva map for a MONICA registered session.

4.2. Data Management and Associated Monitoring Services

As already stated, data, mainly of the time series type, must be centrally stored. For this, a NoSQL-based solution was chosen: MongoDB Community Edition. This type of database, when using a proper schema for time series data, is well-suited for data coming from IoT devices. Moreover, horizontal scaling features like replica sets and sharding provide the necessary capabilities for a scalable and fault-tolerant solution, with authentication and authorization features available at the collection level. The following base collections were saved in MongoDB:

1. MONICA sessions raw data
2. MONICA sessions calibrated data
3. Fixed stations raw data, factory calibrated data and calibrated data using RF
4. Derived stats for all the above collections, populated automatically thanks to change streams feature of MongoDB
5. AQI data derived from data at point 4, thanks to change streams feature and a JS code (executed in NodeJS) to compute the AQI index (with color reference)
6. Stats data by user for the citizen pricing campaign, even this computed using a JS app reacting to events generated by MongoDB change streams

The choice to separate data into different collections was driven by the expected query patterns. The data schema was of the bucket type, with daily or session-based grouping in accordance with MongoDB best practices [41].

As described previously, the entire software solution is made up of many JS applications. Because both the front-end and the back-end use JS as their core, it constitutes what is known as a full-stack JS solution. These applications are monitored using PM2, a commonly used process manager that monitors all the JS applications and restarts them in case of failures. To be notified in case of unhandled exceptions, Sentry [42] (a cloud service for remote error monitoring) was used for both the front-end app (VUE SPA) and the back-end services. Similarly, Uptrends (a website availability monitoring service) was used to monitor the general availability of the web server. The software solution also sees the use of Failban [43] to monitor the NGINX logs and react automatically to handle anomalous traffic towards the server hosting the REST API and the SPA.

4.3. Lessons Learned from Software Integration and Utilization

The software solution/platform developed for Air-Heritage was able to effectively handle the IoT device data from both MONICA and fixed stations, with some real-time functionalities thanks to the change-stream feature available in MongoDB for the replica-set configuration. Nonetheless, some elements could be improved, and although every solution has its pros and cons, the main findings from the experience gathered by the authors can be synthetically summarized below:

- The JSON serialization format could effectively be replaced by Protobuf for transmitting data from the Android/Raspberry Pi device to the cloud/remote server. This change could allow for more efficient data transmission, as Protobuf messages are much smaller in size compared to JSON (from 20% to 80% smaller than equivalent JSON messages) and this reduce system latency [44].
- MQTT is definitely the preferred protocol at OSI layer 4 to use instead of HTTP [45].
- MongoDB was a valid and effective solution. However, current versions of MongoDB (especially starting from v.7 and later) implement TS collections that simplify code development while automatically handling IoT data in an efficient way. Sadly, this introduces some limitations, such as lack of support for change streams widely used within the project, document size (4MB compared to the generic MongoDB document limit of 16MB), and more [46]. These limitations, hopefully, will be removed in future versions of MongoDB. For data coming from IoT devices, other solutions more tailored to TS data could be preferable, such as TimeScaleDB, InfluxDB, QuestDB, to name a few.
- The REST API framework used within the project could be replaced with a solution based on a different language, such as Fiber for Go or FASTAPI in Python. However, a Node.js-based solution is still, in the author's opinion, a valid choice
- To further improve scalability and fault tolerance, it could be useful to revisit the entire solution by employing Kubernetes, with all services deployed using Docker containers and adopting CI/CD practices [47].
- Although the FAIR principles were fulfilled as much as possible, adding an ontology to the data stored in MongoDB could certainly be a very attractive option [48].

5. Effective Logistics Management for Co-Location Campaigns

5.1. Lessons Learned from Software Integration and Utilization

The co-location of sensor nodes with a duly certified reference instrumentations is a prerequisite for progressing to the succeeding phase of node calibration by means of machine learning techniques [49]. It also convey significant information for characterizing sensor response in real world scenarios. The aforementioned step of the Air Heritage project saw a fleet of 30 MONICAs undergo three separate co-location periods in three distinct geographical areas of the Portici city. A mobile regulatory grade analyzer truck by ARPAC (Campania Environmental Protection Agency) was used as reference station in all co-location campaign while the characteristics of the reference air pollution analysers parameters are reported in Table 1.

Table 1. Instrumentation used on board the reference mobile laboratory.

Parameters	Technology	Type	Units
NOX	Chemiluminescence	Thermo Scientific Mod. 42i	µg/m³
CO	Non-dispersive infrared spectroscopy	Teledyne API Mod. T300	mg/m³
O ₃	Ultraviolet photometry	Teledyne API Mod. T400	µg/m³
PM10/PM2.5	Beta-ray attenuation	FAI Mod. SWAM 5a Dual Channel Monitor	µg/m³

The first co-location took place on 2021-01-13 at Portici WaterFront until 2021-03-24, the second co-location period start on 2021-07-04 at Isola Ecologica in Portici up to 2021-10-04 but with some long suspension period due to temporary unavailability of the mobile laboratory. Further details are given in metadata file associated to the publicly available Air-Heritage datalake (see Data Availability Statement paragraph) . Finally, the third co-location took place instead in Scuola Da Vinci Comes in Portici on 2022-01-11 until 2022-04-13.

The co-location campaigns occurred concurrently with the spread of the COVID-19 and the introduction of legislative constraints in Italy due to the state of emergency. These restrictions

effectively constrained the operations of the staff in the assembly and maintenance phases of the entire campaign of co-locations.

5.2. Logistical Challenges and Implemented Solutions

Whether to co-locate all 30 units simultaneously or to divide them into groups for successive periods was one of the first concerns to be addressed. The latter option was selected for two fundamental reasons:

- The placement of the particulate matter instrumentation and other equipment's of the institutional-grade mobile laboratory has resulted in a reduction of the available space on the roof of the vehicle, which has in turn constrained the capacity to accommodate the entire fleet of nodes.
- In order to prevent data loss and collisions between packets during transmission, a maximum of 10 nodes were connected to a single Raspberry Pi, which was employed as a concentrator node for aggregating and forwarding raw data from the nodes to the web server.

In order to guarantee an adequate quantity of data for the training phase, a minimum of three weeks of co-location is required for each individual batch.

Despite the MONICA device's internal rechargeable battery, which is exerted for use in mobile sessions, when colocated it is powered by a 10-Port USB hub when arranged in a fixed configuration. This is due to the fact that the device draws its power supply from the mobile laboratory's main power distribution. Figure 8 illustrates the position of the co-location campaigns on the map and provides a detailed account of the assembly phases.

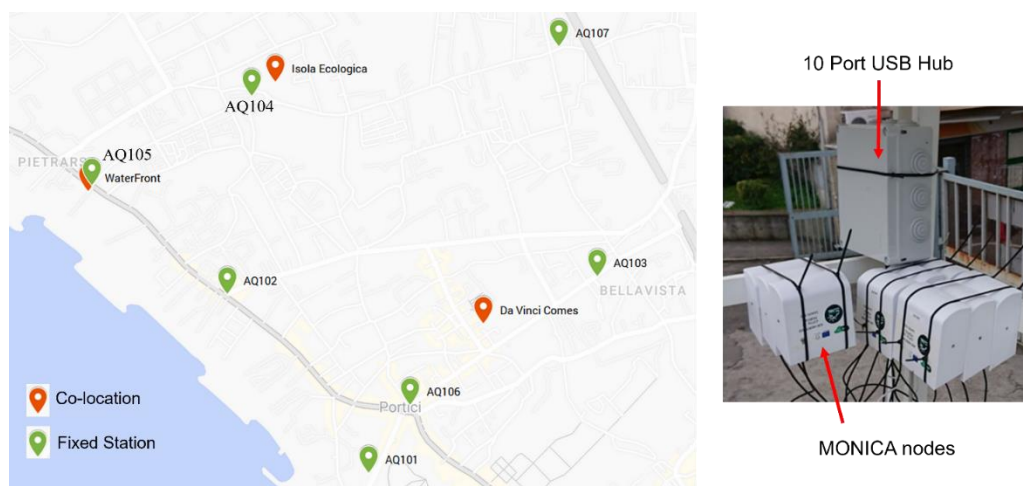


Figure 8. The position of the three co-location campaigns on map performed in Air-Heritage Project and a detail of of the assembly.

The MONICA raw sensors data are captured and transmitted via Bluetooth Low Energy (BLE) to a Raspberry Pi Mod. 3B+ based datasink with Raspbian OS providing for WAN connectivity services through a mobile router wi-fi TP-Link M7650. Data acquisition is managed by a Python script running on the Raspberry Pi. At remote side, an ad-hoc IoT backend architecture relying on a contained NodeJS REST APIs server and MongoDB provides data inception, device management, storage and data pre-processing (see Figure 9).

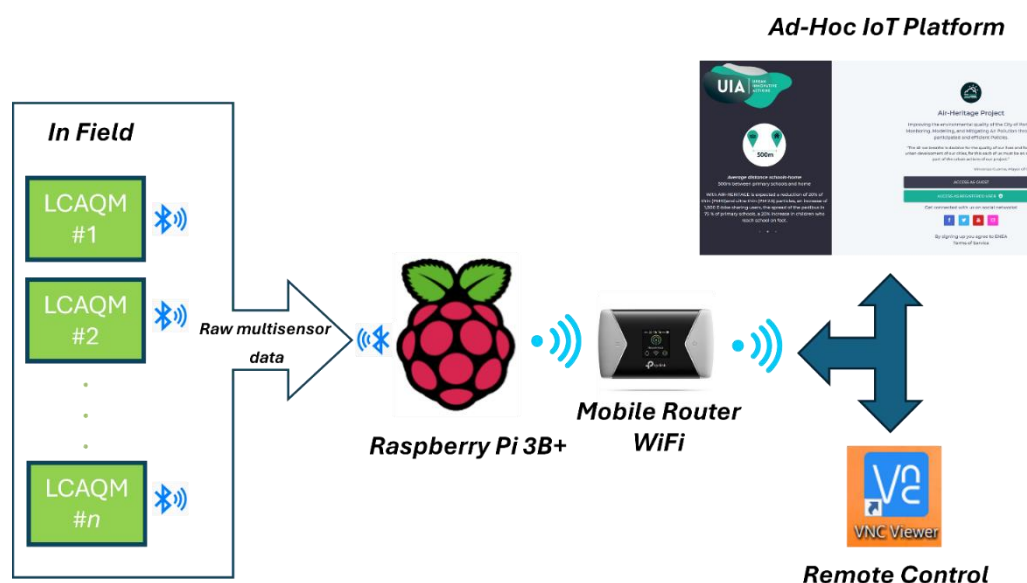


Figure 9. Scheme of IoT architecture in stationary setup.

The integration of a Wi-Fi router into the IoT infrastructure design provides the additional benefit of enabling remote control of data transmission. This is achieved by accessing the Raspberry Pi as a remote desktop with the RealVNC® application. This enables the remote monitoring of data transmission, allowing for the assessment of device functionality and the identification of instances of non-transmission. In the event that on-site verification is required, it can be scheduled. It is crucial to acknowledge that any issues with the reference mobile vehicle power supply may impact on the entire network of the co-located devices. In the event of a power shortage, the internal battery of the Wi-Fi router will be discharged, thereby rendering remote access impossible.

During the second co-location campaign, which spanned from 2021 July 4th to 2021 October 4th at Isola Ecologica, the unavailability of the ARPAC mobile regulatory-grade truck, due to emergency relocation, necessitated the co-location of a number of nodes exceeding 10. This was done in order to expedite the process in anticipation of future Citizen Science campaigns. In this particular situation, we elected to utilise two Raspberry Pi, apportioning half of the nodes to each individual Raspberry.

The latter highlights the importance and the difficulties related to the access to reference stations which, in our case, was made possible thanks to the collaborative efforts of ARPAC. Direct access may also be impossible in several settings due to the unavailability of locally deployed reference stations or the lack of resources needed to guarantee this access for research personnel to the regulatory grade stations. It is worth to remember, that these, in turn, are designed and operated for different goals and implement strict regulatory framework to keep being source of authoritative data for the community.

5.3. Recommendations for Future Co-Location Efforts

The IoT infrastructure adopted solution has been demonstrated to be an effective and dependable approach for managing the data load of a subset of ten nodes. It is evident that the solution utilising Raspberry Pi is a cost-effective and open-source option, particularly when combined with Python scripts.

Any effort should be made to allow for closeby co-location between all nodes and the reference analyzers. In case of a high number of LCAQMS with significant dimensions this may represent an issue but accurate colocation planning after a site inspection may allow for optimal trade off (see Figure 10).

The decision to divide the fleet into three batches per co-location was largely a practical necessity in our case. However, the management of a limited number of devices has proven to be an effective approach. This method allows for the periodic remote monitoring of node transmission status, which

is a valuable tool for maintaining operational efficiency. Currently, this activity is performed manually and on a periodic basis. It is recommended that this process be automated in the future by developing a series of Python scripts that monitor the data transmission status and send an alert in the event of prolonged lack of received data packets. Such an approach would facilitate timely intervention to resolve the problem, which could be of an electrical nature, a lack of power supply, or a variety of issues specific to the individual node. Furthermore, the management of a limited number of co-located nodes offers the additional benefit of facilitating maintenance activities.

Finally, we have attempted to utilise batching in order to assess the efficacy of novel calibration methodologies, including global calibration [50]. This methodology was developed with the specific objective of reducing co-location and calibration costs through the utilisation of a virtual calibration model that incorporates the inherent variability of individual sensors. The findings indicate that co-locating a minimum of five nodes is sufficient to apply the virtual model to the remaining nodes for low-cost particulate matter sensors [51]. This approach is undoubtedly beneficial in the context of large-scale deployments. Further research is required to validate this approach on large deployments and to ascertain whether an extension to gas sensors is also feasible [52].

6. Calibration and Data Management

Most Low cost Air Quality sensors needs a calibration law which translate their raw response values, usually encoded in electrical variables, in actual concentration data. Typical example are MOX which encode concentrations in resistivity or Electrochemical sensors which encode concentrations in electrical potentials, or QMB which response are encoded in frequency shifts [53]. Unfortunately, along with their undoubted advantages, they suffer from several operational issue which hinder calibrated values accuracy, these effects are particularly significant when field deployed. Among these, fabrication variance, which imply sensing unit characteristics variance, lack of stability, which make these characteristics to vary in time, environmental and non target gases sensitivity which make their response vary with those factors intensities, are the most challenging. Their combined effect requires extensive and/or frequent recalibration actions in time or after relocation [54–56]. These considerations pointed us to proceed to develop a peculiar though literature based, calibration strategy so to keep a balance between due efforts and accuracy with the latter as a priority. These, together with devices batch size, availability for actual deployments, and the need for reference data constituted the operational and functional requirements which lead us to devise a seasonal recalibration protocol exploiting field calibration as the preferred and most accurate methodology [57]. Whenever available, colocation data provide in facts for the best reference data to achieve a data driven calibration.

According to the plan described in previous chapter and detailed in Table 2, all fixed and mobile devices have been collocated with a reference analyzer at least for three weeks in three different seasons and locations within the city of Portici, the target city of the Project. Actual operative deployment occurred between 2 consecutive calibration periods. Figures 11–13 actually shows some of the operative deployment preparation phase including pre-charging for mobile devices and the deployment of one of the fixed units. All raw data recorded during the colocation periods have been captured along with reference analyzers readings provided by regulatory grade mobile monitoring station. Time reconciliated data series have been collected in an ad-hoc datalake comprising both fixed stations and mobile stations colocation data in order to support the selection, development and test of calibration algorithms.

Table 2. Colocation scheme of the mobile MONICA devices. The schema encompasses 3 Periods in 3 different seasons and locations within the urban area of Portici (Naples).

Period 1 (2021)			Period 2 (2021)			Period 3 (2022)			
#	Batch 1	Batch 2	Batch 3	Batch 1	Batch 2	Batch 3	Batch 1	Batch 2	Batch 3
	Jan, 13th 15:00 ->	Feb, 5th12:00- > Mar	Mar, 2nd 14:00 ->	Jul, 4th 00:00 ->	Aug, 24th 11:00	Sep14 th 1015 ->	Nov 1st, 00:00 ->	Feb 9th,00:00 > March	March4t h,00:00

	Feb, 5 12:00;	2nd, 10:00	Mar 24th 10:00	Jul 19th 23:59;	>Sep,14 th8:40;	Oct,4th 9:20	Feb, 2nd, 23:59	3rd, 23:59	>Apr,13t h,23:59
1	337	325	324	332	327	324	324	327	325
2	339	326	330	340	330	325	326	328	356
3	344	327	334	349	331	326	329	349	353
4	345	329	335	350	333	329	334	355	350
5	349	331	343	353	334	343	339	347	335
6	353	332	350	356	335	—	338	345	330
7	355	333	351	360	337	—	344	331	332
8	356	340	352	362	339	—	361	333	341
9	360	341	362	—	341	—	—	337	351
10	361	364	363	—	344	—	—	343	—
11	—	—	—	—	345	—	—	—	—
12	—	—	—	—	351	—	—	—	—
13	—	—	—	—	355	—	—	—	—
14	—	—	—	—	363	—	—	—	—

Generally, AQ multisensory calibration algorithms use raw sensors data together with environmental variables, so to correct their interferences and return concentration estimations as an output. Formally, let $C = (C_1, C_2, \dots, C_n)$ the true concentrations vector and $X = (X_1, X_2, \dots, X_n)$ the input raw sensors readings matrix, calibration stage aims to derive a function f , such that,

$$C = f(X) \mp \varepsilon \quad (1)$$

with ε being the residual calibration error.

As we have seen before, raw gas sensors data encodes forcers information as analog voltages, frequency shifts, or IR spectras, whilst PM sensors either use particle count as raw data or provide a vendor based calibrated output which can be used as “raw” input data. As we introduced above, response forcers may include target gas and other gaseous interferences as well as environmental interferences. X vector may turn in a matrix if including past records of the same variables and/or past target concentrations as stored in tapped delay lines (lagged values) for deriving dynamic and/or recurrent models, respectively. In our case, the input variable array X included raw data from the chosen electrochemical sensors (working and auxiliary electrode voltage) for target gas and interferences, plus environmental interferences. For NO₂, this is expressed as:

$$X = \begin{bmatrix} WE_{NO_2} \\ AE_{NO_2} \\ T \end{bmatrix} \quad (2)$$

where WE_{NO_2} (in mV) and AE_{NO_2} (in mV) are the output of Working and Auxiliary Electrode, respectively, of NO₂ sensor, T is the Temperature (in °C). In facts temperature is reported as one of the most significant interferences for electrochemical gas sensors. Alphasense itself advised to correct concentration estimations according to temperature values.

For CO gas concentration estimation, we used CO sensors WE and AE voltages, plus temperature readings. Hence, X is instead given by:

$$X = \begin{bmatrix} WE_{CO} \\ AE_{CO} \\ T \end{bmatrix} \quad (3)$$

Due to well known interference of NO₂ concentrations on the chosen sensors, O₃ concentration was estimated using sensor readings coming from NO₂ and O₃ sensors plus temperature readings:

$$X = \begin{bmatrix} WE_{O_3} \\ AE_{O_3} \\ WE_{NO_2} \\ AE_{NO_2} \\ T \end{bmatrix} \quad (4)$$

Conversely, Humidity served as primary correction factor in the MLR model formulation for PM sensors, for the MONCIA devices we hence had:

$$X = \begin{bmatrix} Conc_{PM2.5/PM10} \\ RH \end{bmatrix} \quad (3)$$

Note that vendor estimated size fractioned concentration for the targeted pollutant was used as raw sensor for PMS7003 sensors calibration along with Relative Humidity readings.

According to literature and to our field deployments experience, 3 weeks colocation data are sufficient to obtain optimal short term accuracy performances [56,58]. Regarding the model choice, multilinear regression is actually providing a good balance between short term precision and accuracy metrics while showing interesting generalization properties for mid term deployments [59]. Comparatively, Random Forest model among non linear machine learning approaches is usually referred as providing best accuracy for short term deployment but fail to provide accurate concentrations readings when operating outside the input ranges within which it has been trained. MLR has been hence been chosen as a data driven model for providing calibration for both Gas and Particulate sensors for mobile and fixed stations according to

$$y = aX + b \quad (5)$$

where b is the intercept and a are the least square optimal coefficients.

Summarizing, to ensure the highest accuracy during long-term deployment, the AirHeritage project has implemented an edge computing, field calibration, data-driven, MLR model based calibration scheme for 30 mobile analyzers, with periodical season based recalibration. Accuracy figures estimations for the MONICA devices, obtained with 2 consecutive weeks long calibration set during period 1, are reported Table 3: Crossvalidated MLR performance metrics averaged for the 30 MONICA devices (2 Consecutive Weeks Training length, 1 Week Test Length, 2 temporal stratifications over 3 weeks) during the first colocation period. Negative estimations are reverted to null concentrations. It is important to stress the importance to estimate performance using consecutive temporally stratified training and test set, avoiding any data leaks so to simulate real deployment conditions thus obtaining realistic performance indexes estimations. In terms of coefficient of determination, depending on the pollutant specie we obtained values which span from 0.43 of PM10 estimations to 0.88 for Ozone concentration estimations. Low values obtained for PM10 can be explained with the low sensitivity of the specific sensor to that size particulate fraction as it is well known that most of low cost particulate sensors are optically tuned to express maximum sensitivity to PM_{2.5} fraction. Performance of CO sensor are explained with the low concentrations recorded during the specific colocation period (see Figures 14 and 15).

Table 3. Crossvalidated MLR performance metrics averaged for the 30 MONICA devices (2 Consecutive Weeks Training length, 1 Week Test Length, 2 temporal stratifications over 3 weeks) during the first colocation period. Negative estimations are reverted to null concentrations.

		Train	Test	MAE	R ²	RMSE	NRMSE	MAPE
		#Hrs	#Hrs	µg/m ³	N/A	µg/m ³	µg/m ³	N/A
O3	AVG	322.7	207.3	7.46	0.86	9.59	0.35	0.08
	STD	25.6	35.3	1.92	0.10	3.00	0.12	0.02
	MDN	330	203.5	6.99	0.88	9.32	0.33	0.08
NO2	AVG	322.7	207.3	7.92	0.83	10.32	0.41	0.08
	STD	25.62	35.4	2.41	0.08	3.10	0.09	0.02
	MDN	330	203.5	6.95	0.86	9.07	0.38	0.07
CO	AVG	322.7	207.3	190	0.33	260	0.74	0.12
	STD	25.62	35.4	80	1.02	0.12	0.34	0.04
	MDN	330	203.5	190	0.51	240	0.69	0.11
PM25	AVG	323.5	195.5	5.49	0.75	7.43	0.49	0.07
	STD	25.8	35.1	1.12	0.11	1.38	0.11	0.02
	MDN	323.5	195.5	5.13	0.75	7.32	0.50	0.08
PM10	AVG	321.9	206.7	13.07	0.42	20.84	0.75	0.10
	STD	25.4	34.8	4.83	0.19	12.94	0.12	0.03
	MDN	329.5	202.5	11.86	0.43	14.10	0.76	0.10

The analysis of performance variance showed that in this short term deployments, with the exception of short term transients, we did not record performance extremes with all devices showing consistent performance ratings (see Table 3). Just as an example, R^2 ratings showed a minimum value of 0.72 and a maximum value of 0.93. Since actual concentration range is a significant factor influencing performances we also report that MAPE readings was comprised between a minimum of 4.4% and a maximum of 9.4%, in the latter case the device kept an R^2 of 0.78.

Unfortunately, our multiseasonal recordings confirmed that ab-initio field calibration performances show a significant performance drop which requires re-calibration at least on a seasonal basis, which we actually performed. As an example R^2 figures for PM2.5 dropped to an average 0.26, as we reported in [58].

We remind that calibration function was actually executed on a smartphone, enabling immediate and localized feedback to the user while they move across the city which resulted in a really appreciated feature. For doing so, we decided to temporarily store Calibration parameters in the Android app resource files, where they are retrieved at the start of each recording session. This allows for updates without needing to rewrite the device firmware and the need for stable connectivity. However, as cellular connectivity has reached an operational maturity allowing for stable connection almost everywhere, we chose, for the next version of the MONICA ecosystem, to provide an upgrade to store calibration parameters on centralized inception and storage facilities, eliminating the need for app updates. This will also make easier to constantly upgrade calibration parameters if remote calibration strategy will be in place.

Lesson learnt:

Three weeks colocation duration showed appropriate to build a sufficient dataset covering the intra-seasonal concentrations and environmental conditions variance, thus allowing for optimal short term performances [57].

Periodic Field calibration by colocation quickly become unfeasible for the number of devices usually involved in recent projects / commercial deployments which is approaching several tenths or hundreds of nodes [14] while could be still effective for accuracy driven short term experimental deployments [59,60]. Repeated removal and deployments procedures, represents the main cost sources along with the costly access to reference colocation data. They also represent a significant risk for multisensory devices damage.

The use of universal calibration [50] or calibration transfer approaches have been explored as a way to reduce calibration efforts limiting the number of devices to be recalibrates. However, apart from PM sensors which show a limited fabrication variance, our results for electrochemical or MOX gas sensors are still controversial. Eventually if sensors drift kicks in then the accuracy of all calibration methods and models will fade without upgrades.

Remote calibration or periodic in place colocation with gold nodes can provide reference data for reducing drift impacts [61]. Solution has to be chosen according to different parameters including the availability of close enough or correlated/statistically reconciliable source of data in case of remote calibration and or logistic burden for gold nodes opportunistic or routine recalibration [62].

Dynamic calibration with recurrent architectures (say RNNs) could be needed to limit the impact of sensor response time on pollution mapping while on the move. In our case the limited speed of pedestrian citizen which represented the backbone of AirHeritage citizen sciences campaigns and the sensors T_{90} values (<1min) made us confident to resort to static calibration.

Since local, real time, assessment of pollutant exposure meet the citizens favour, we advise for inclusion of such a feature in any portable device. As such the local availability of the calibration model will be important to guarantee this function in real time without resorting to cloud and avoiding the relative costs. Thanks to general availability of cellulare WAN connectivity, we advise to download model updates at each measurement session start.

KPI:

Average Sample loss rate (hours) < 15%

Accuracy performance metrics estimations reported median MAPE of less than 10% for all pollutants while coefficient of determination varied from 0.43 fro PM10 to 0.88 for Ozone.



Figure 10. The 7 fixed stations as deployed nearby the reference mobile station during calibration data gathering (colocation periods).



Figure 11. MONICA devices as ready to be shipped to citizens.

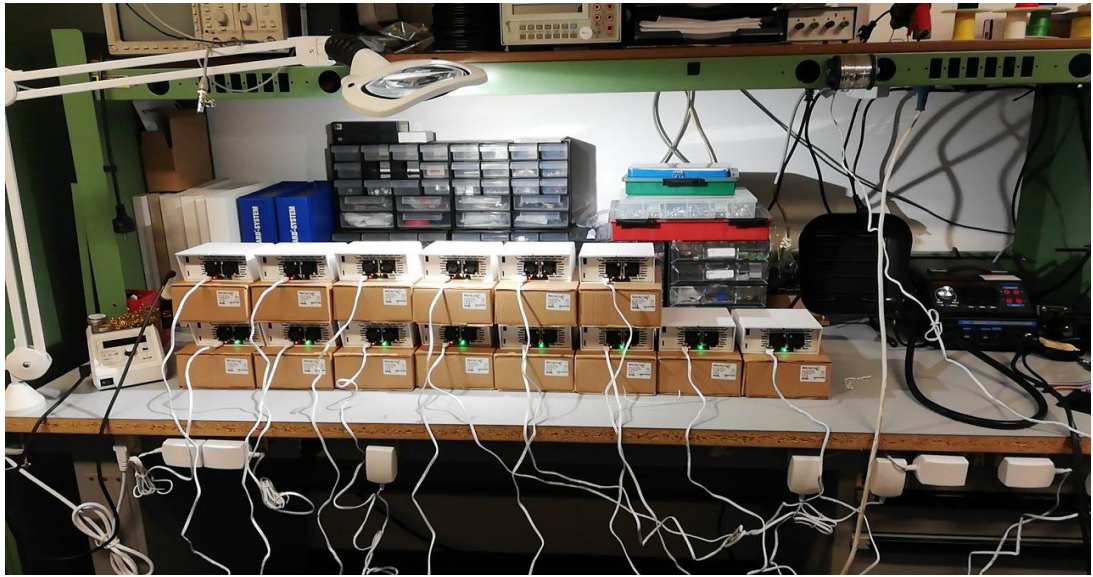


Figure 12. Precharging MONICA devices.



Figure 13. Actual deployment of AirHeritage fixed stations.

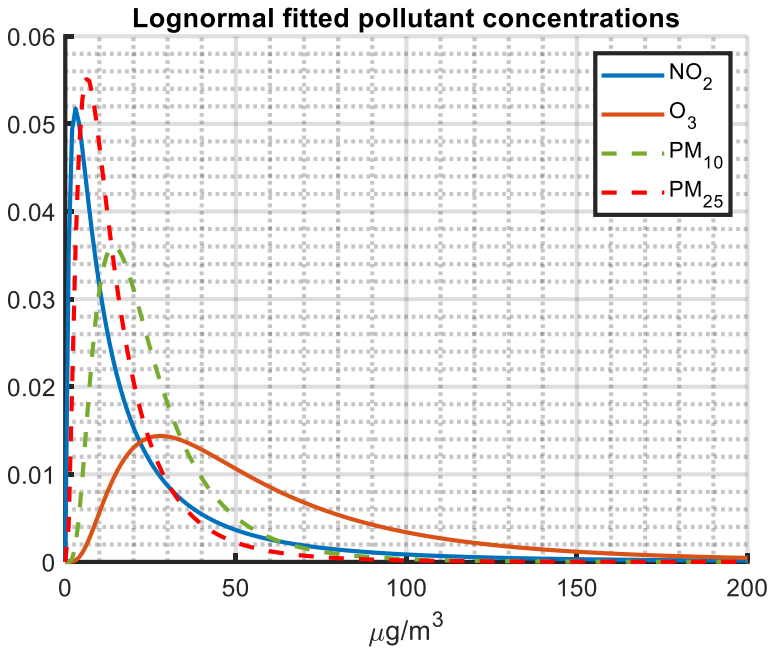


Figure 14. Lognormal fitted pollutant concentrations as recorded in the first colocation period.

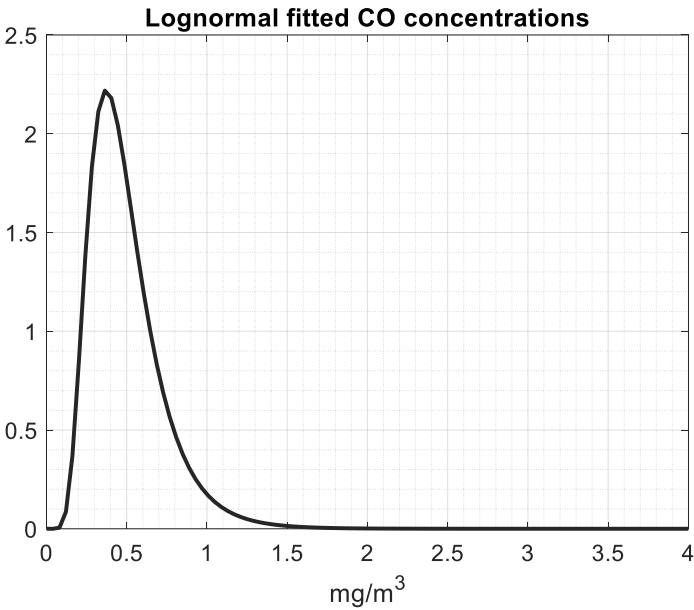


Figure 15. Lognormal fitted concentrations of CO as recorded during the first colocation period.

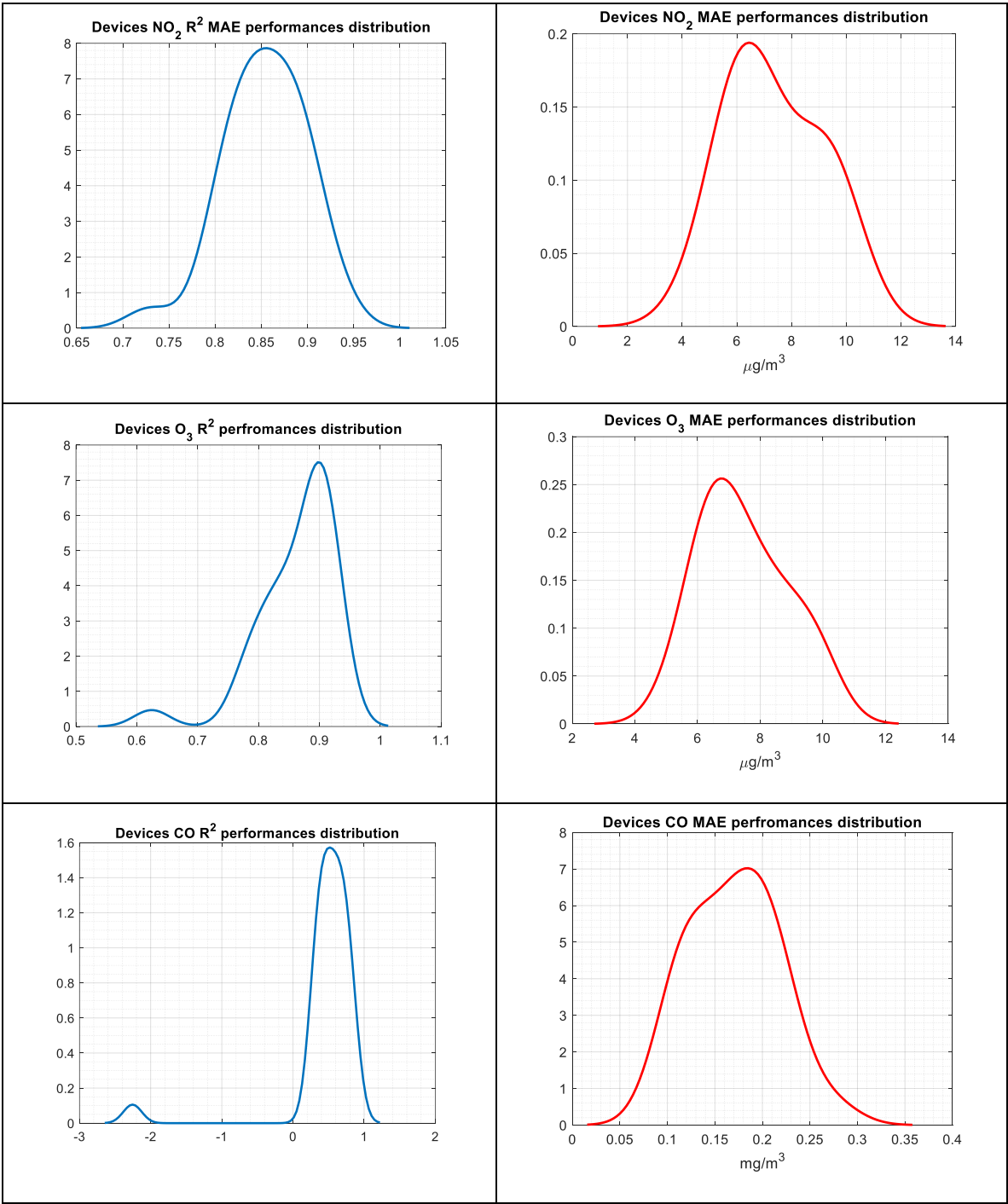


Figure 16. Distribution, across the 30 Monica devices, of R² (first coloumn) and MAE (second coloumn) short term performance values for NO₂ (first row), O₃ (second row) and CO (third row) as estimated by MLR based data driven calibration in deployment period 1. The distributions appear to be skewed by a few outliers. Performed checks show anomalus low performance are due to transients in raw sensor responses when they were first switched on.

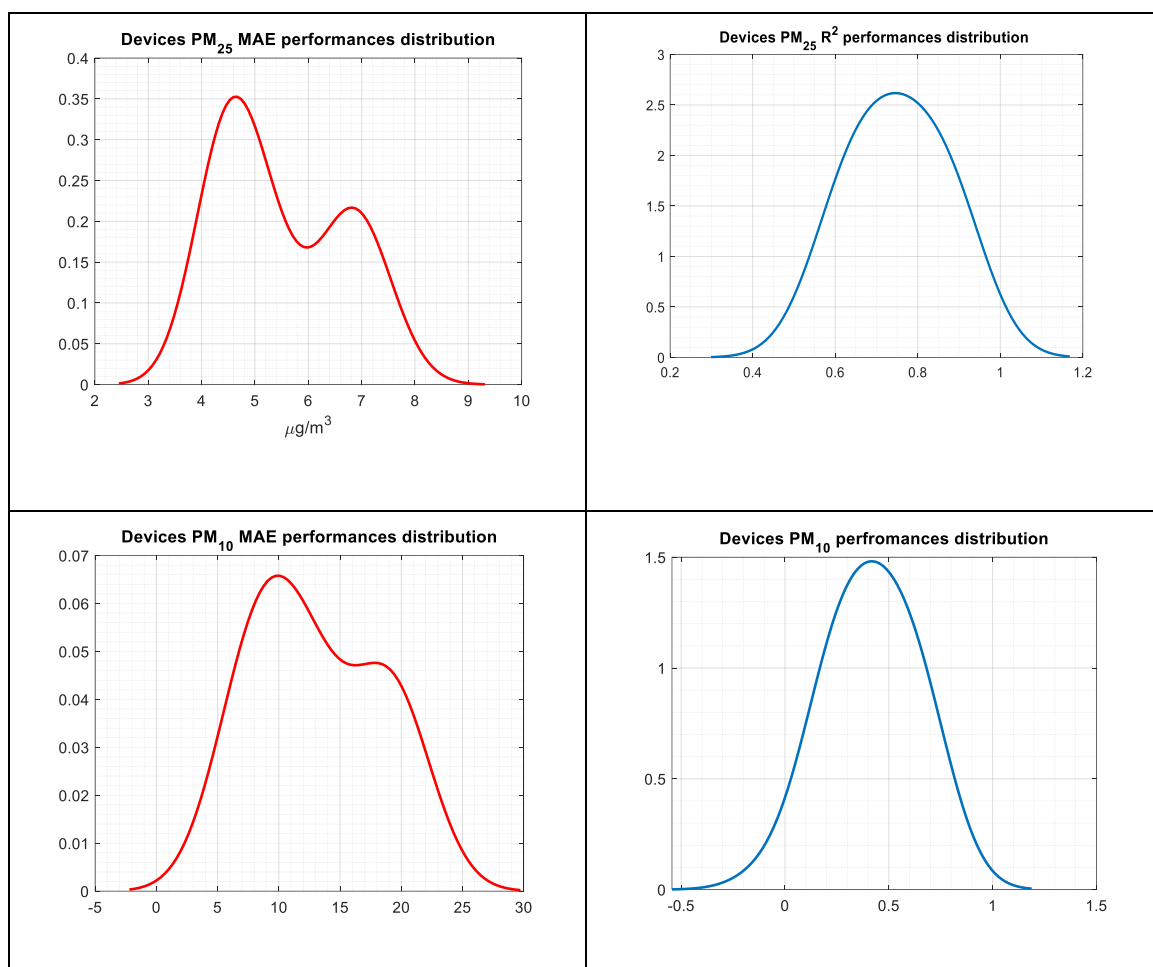


Figure 17. MAE (1st coloumn) and R^2 (2nd coloumn) and short term performance for $PM_{2.5}$ (first row) and PM_{10} (second row) as estimated by MLR based data driven calibration in deployment period 1 , across the 30 Monica devices.

7. Citizen Engagement Through the MONICA App

7.1. Introduction and Role of the MONICA App in Citizen Engagement

As discussed previously, the primary objective of the co-location phase is to calibrate the sensor nodes, and, just in the case of fixed stations, to monitor the distributions of pollutants in a single geographical point. However, to observe how pollutants are spatially distributed over time, mobile campaigns must be carried out. In this instance, the collaboration of volunteer citizens was enlisted for the implementation of mobile campaigns within the city of Portici. They literally "strolled" with MONICA through the city. The local associations, which were involved by the Air-Heritage project partners, played a pivotal role in facilitating citizen involvement.

In order to guarantee the accurate transmission and storage of data collected by sensor nodes during citizen science campaigns, a dedicated smartphone application was developed. In essence, the user's smartphone serves as a gateway between the sensor node and the web.

The MONICA App is currently developed exclusively for Android smartphones using Android Studio. In order to ensure consistency with the developed IoT platform, Auth0 is employed for authentication services. Regarding distribution, we opted for third-party distribution, thereby making the .apk file accessible directly on a proper web page of the developed IoT platform (Figure 18).

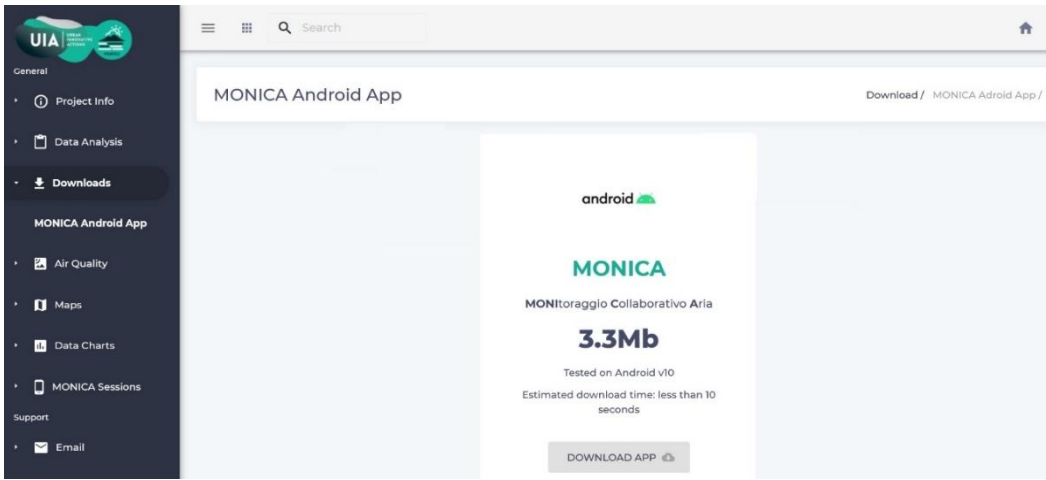


Figure 18. The web page of the IoT Platform for downloading the MONICA app .apk file.

Of course, not all end-users were previously familiar with the installation of third-party applications on their mobile. Consequently, training events were organised to address this issue, which also included training in the use of the app and in the use of the web page for viewing the data of the sessions carried out. The user is able to access and view the data of their own sessions both from the back-end and on the appropriate screens of the app, as illustrated in Figure 19

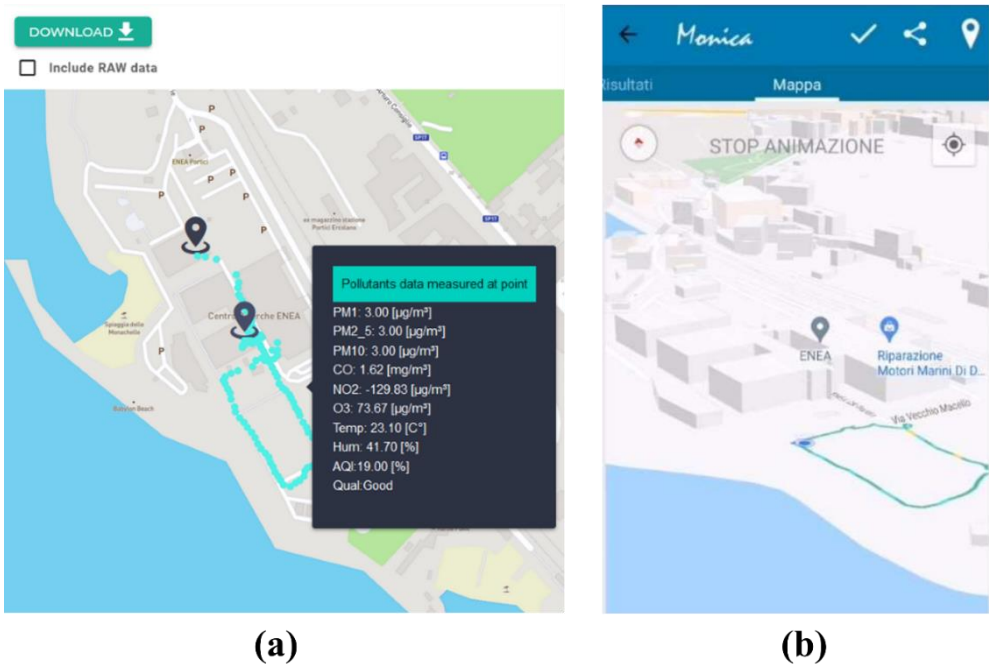


Figure 19. (a) An illustrative example of a user session as displayed on the webpage, accompanied by an indication of the location and the level of pollutants. (b) An illustrative example of a user session as displayed on the MONICA app.

7.2. How the Monica app works

The MONICA app is responsible for receiving sensor data via Bluetooth Low Energy (BLE), returning the concentration values relating to individual pollutants (calibrated data) and sending them together with the raw sensor data to the back-end server. Figure 20 provides a schematic representation of the data flow.

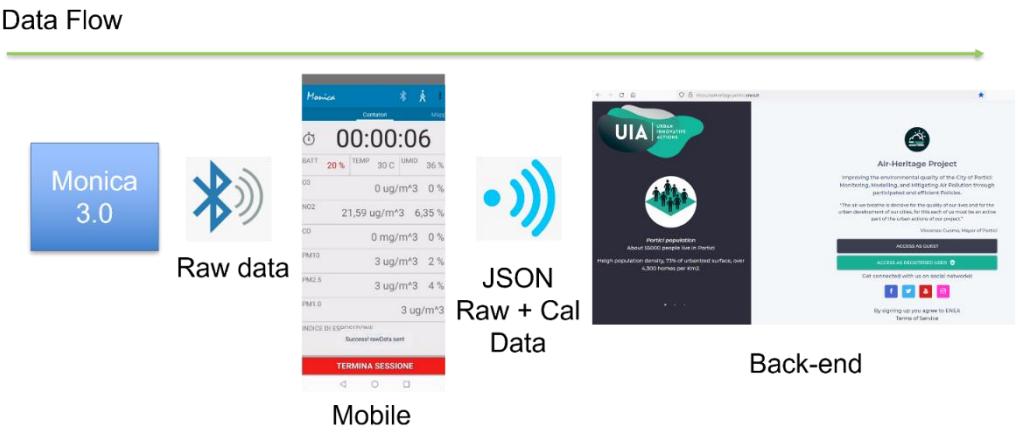


Figure 20. A schematic representation of the data flow in a mobile application scenario.

The unprocessed data from each MONICA sensor node are encapsulated in a JSON format and transmitted via BLE to the user's smartphone. The `BLEService` class, as defined within the code, provides connection services with remote devices, intercepts and filters BLE communications, as well as routes these communications to the intended recipients. This enables the other components of the app to request actions from the `BLEService`. Unlike the location notification and update service, the `BLEService` is declared as an external background service. The calibration parameters, i.e. the MLR coefficients, are stored in a resource file called `LinearCalibration` which is updated after each co-location with the new appropriate coefficients. Figure 21 delineates the sequence of operations comprising the comprehensive workflow, from the first co-location until the completion of the first citizen science campaign. The process is naturally repeated for each subsequent co-location.

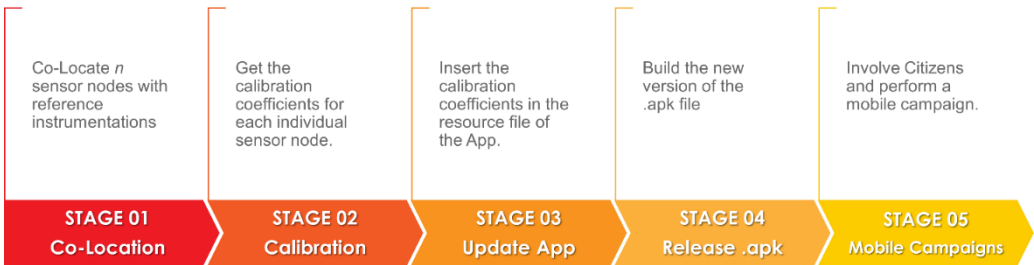


Figure 21. The workflow performed in the Air-Heritage project.

An ad-hoc developed function denominated `BuildJSON` is responsible for the generation of a single JSON, which contains both the raw and calibrated data. Additionally, this function is responsible for creating the necessary sockets to send the raw data into the MongoDB raw data collection and the calibrated data into the MongoDB calibrated data collection.

The MONICA device was designed as a wearable device, therefore it is not equipped with an internal memory card for storing the data of each individual session. Instead, the data must be sent to a remote server. In the event of a lack of network connectivity in specific locations during the monitoring session, there is a possibility of data loss. In order to prevent this situation, the app is capable to buffer the session data for two minutes. In the event that the network becomes available during this period, the buffered data is transmitted to the server in conjunction with the newly acquired data. In the event that the buffer is full and the network is not yet available, the session is closed proactively. In order to inform the user of the occurrence of such an event, we decided to

display a "Toast message" on the smartphone screen. This message indicates whether the raw and calibrated data has been successfully transmitted to the server. Consequently, the user is able to take the necessary action to terminate the session and/or proceed to an alternative location where the network is accessible.

7.3. Feedback from the Citizen science campaigns

During the testing phase of the app and during the mobility campaigns, the app code was updated on a regular basis in order to ensure compatibility with the latest Android releases. This process was necessary in order to reach a stable version that would guarantee adequate compatibility. In order to familiarise users with the process of downloading and installing third-party apps on their smartphones, both virtual and in-person meetings have been provided. The management of the training phase of a large number of users and the concomitant updates to Android releases proved to be a significant challenge, requiring a significant expenditure of both time and human resources. The burden would have been significantly increased if we had decided to develop also an Apple version of the app and distribute it on their respective marketplaces. Nevertheless, this experience has led us to conclude that future app development should adopt multi-platform programming languages such as Flutter™ and over and above distribute the app through their respective app stores.

It must be observed that the involvement of the general public in citizen science campaigns has been confined to a limited circle of volunteers from the participating associations and their acquaintances, largely engaged through word-of-mouth communication. The number of individuals from outside voluntary organisations who participated was negligible, despite extensive social media campaigns to advertise the campaigns. It would be beneficial to identify alternative methods of public engagement for future projects.

8. Impact of Spatial Analysis of Citizen-Generated Data

Capturing the local variability in air pollutant concentrations within the cities is a key issue to assess the urban air quality at high spatial resolution as well as to evaluate the citizen exposure to high air pollution levels [10]. To address this issue within the project, a pervasive urban air quality monitoring using hybrid deployments of sensor networks has been investigated.

As shown in literature, the high spatial variability of urban air pollution is not usually captured by regulatory monitoring networks being sparsely distributed within the city. As first task, the deployment of fixed low-cost sensors networks has been designed and developed across the pilot City of Portici.

Using networks of stationary low costs sensors within a city, neighborhood-level differences as well as near specific sources (e.g., roadways) can be captured and pollutant spatial variations at fine (sub-km) length scales can be quantified. At this scope, the relevant question of the "optimal sampling design" has been investigated. Thus, optimal sites for deploying stationary low-cost sensors have been identified aimed at a spatially dense air pollution monitoring on the basis of which more reliable air quality forecasts across the city can be obtained as well as population pollutant exposures can be evaluated more accurately.

To do that, a spatial multicriteria analysis (GIS-Multi-Criteria Decision Assessment) has been developed for selecting suitable sites for low-cost multi-sensor traffic-orientated stations across the city [63].

The street canyon effects and the traffic-related emissions were selected as suitability criteria intending to represent the impact of local sources and land use on measured pollutant concentrations. In fact, street canyon effects, generated by the ratio between the buildings height and the streets width, can modify concentration patterns near roadways. The traffic-related emissions contribute to background pollutant concentrations but they are also the principal source of local variation in the concentrations of urban air pollutants. Higher pollutant concentrations are in high traffic emissions areas.

High- and low- traffic emission zones and hybrid zones were identified by a spatial hot spot analysis. All selected criteria were performed as geographical variables and combined in a weighted spatial overlay. For the pilot city of Portici, a map of suitable sites was generated for deploying low-cost multi-sensor traffic-orientated stations for monitoring NO_x and PM_{2.5} concentrations and quantifying their spatial variability until urban streets (Figure 22).

Among the main results of this approach, we can list:

- Low-cost monitoring stations located at the eligible sites so identified can convey information on areas on which space variability is significant, providing that informative content which is actually lacking for both regulatory monitoring networks and modelling based approaches for air quality mapping.
- In addition to the local spatial variability, the temporal variability of air pollutant concentrations has to be taken into account for obtaining more reliable urban air quality scenarios.
- One of the possible limitations to the use of the proposed spatial analysis is its reliance on data. Data could be difficult to obtain such as vehicular flow (simulated or measured) as well as the street canyon effects. In these cases, the use of proxy data could partially solve the issue.

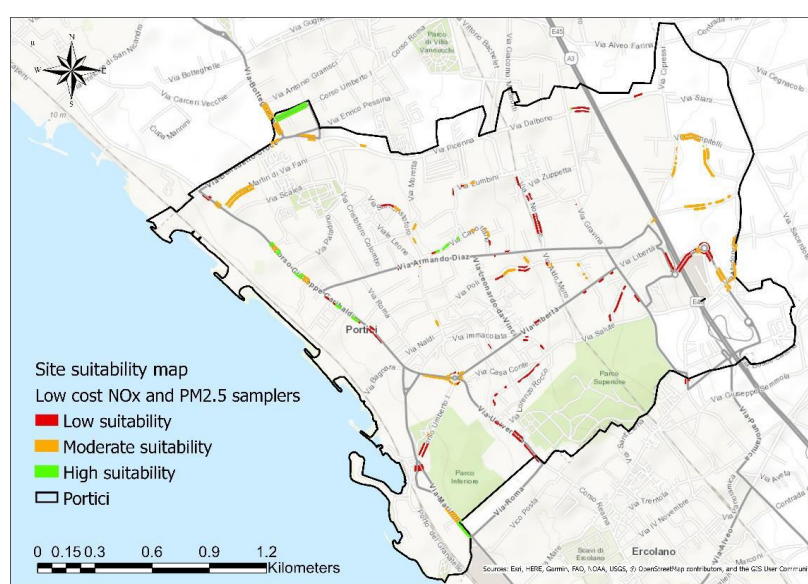


Figure 22. Site suitability map for networks of low-cost traffic-orientated stations for air pollutants monitoring across the city of Portici.

Also, air quality data gathered by stationary low-cost (multi-) sensor networks as well as mobile and opportunistic monitoring campaigns across urban areas can be fundamental in data assimilation approaches aimed at including the variability patterns of air pollutant concentrations that conventional chemical transport models or land use regression models fail to adequately represent.

In that regard, we have investigated the capability of an operational air dispersion model at district scale, such as SIRANE in identifying the local spatio-temporal variability of pollutant concentrations within an urban district, with respect to mobile monitoring campaigns, using the low-cost sensor devices – MONICA [64,65]. The campaign consisted of 6 runs, along a fixed monitoring route, repeated at two different days and different times on one day (Figure 23). By comparing the mean concentrations of CO and NO₂ pollutants measured by MONICA devices and those simulated by SIRANE along the urban streets, we have shown that the recorded concentrations appear exceeding the simulated ones by a factor of 3 and 2 for CO and for NO₂, respectively. Furthermore, for both the pollutants, this factor is higher within the street canyons than in open roads. However, the data recorded and the simulated pollutant concentrations show patterns that broadly correspond, and also the peaks observed in the mobile profiles appear in the simulated profiles, though smoothed (Figure 24). The outcomes highlighted that there are differences between measured and estimated pollutant concentrations due to several factors. The operational conditions are one of these. MONICA

measures pollutants concentrations at about 1 m height and at high temporal resolution, capturing also concentration peaks due to events that occurred during the monitoring run such as a closely passing car or bus, or walking in the emission plume right behind a vehicle, or even idling vehicles due to local traffic congestion. SIRANE evaluates hourly concentrations, spatially averaged in each street segment.

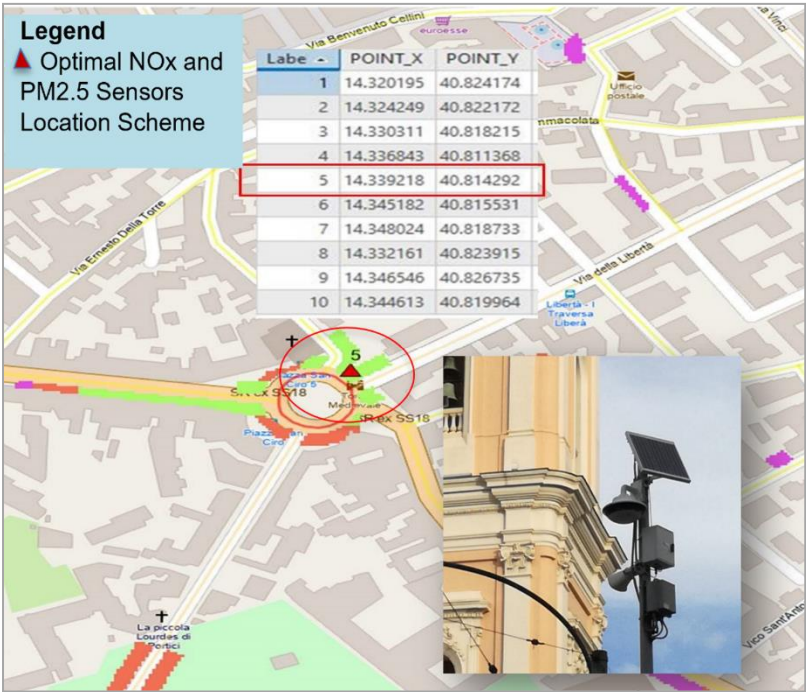


Figure 23. A view of the optimal location (evaluated highly suitable) of Piazza San Piero (City of Portici) for the NOx and PM2.5 sensors to be installed across the city.

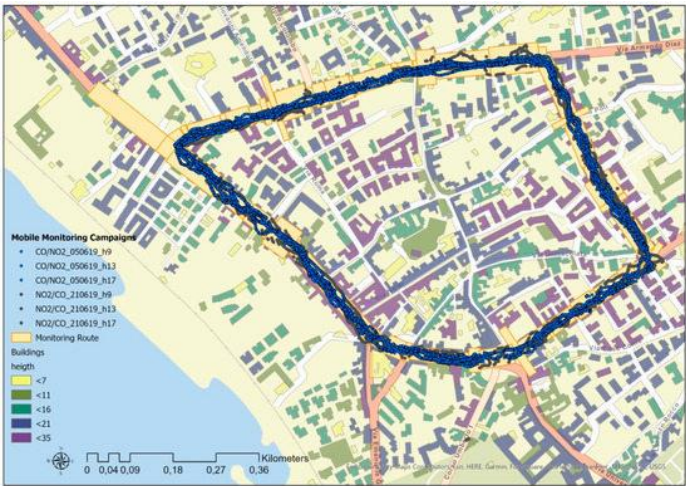


Figure 24. Maps of the mobile monitoring campaigns along the selected monitoring route.

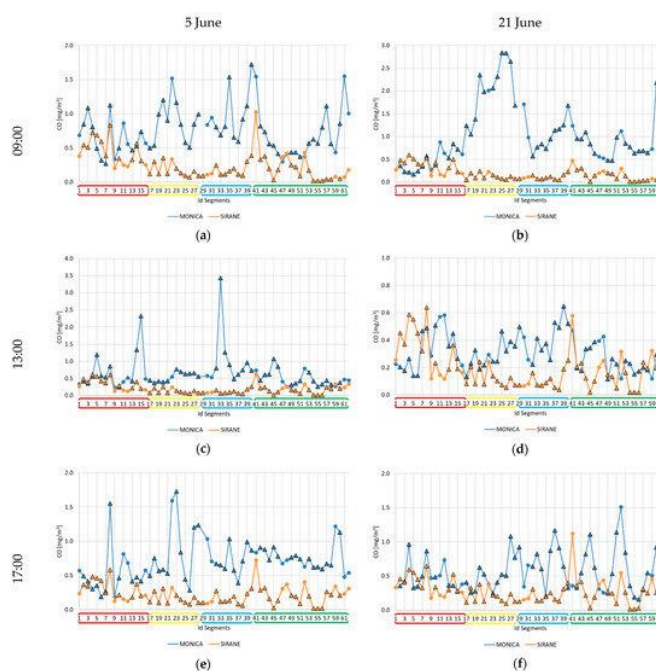


Figure 25. Comparison between MONICA (blue line) and SIRANE (orange line), for CO pollutant on 5 June. Triangles are street canyons and circle open roads. The ID receptors are grouped by monitoring road segments.

The further task was to explore the potential of participatory monitoring campaigns for mapping the exposure to air pollution in the urban environment at a high spatial resolution. This monitoring approach made use of citizens' common daily routines to move low-cost air quality measurement devices as MONICA around the city. It is a promising approach based on un-structured setup to collect large datasets that give useful additional information at a reasonable cost compared to classical data collection methods. But, depending on the set-up of the data collection, such new data can lead to new challenges in data processing and interpretation. At this regard, a statistical analysis method as spatial binning was applied in an original way for exploring the distribution of the large dataset collected by the citizens carrying MONICA devices during their daily walking routes around the city results in unstructured measurements without distinct patterns in space or time.

The collected measurements were grouped in a summer (from April to July) campaign and a winter (from October to December) campaign on most of urban roads. About 46.000 raw 5 s measurements of CO, NO₂, O₃, PM_{2.5} e PM₁₀ concentrations were recorded during the summer campaign and about 100.000 raw 5 s measurements of the same pollutants during the winter campaign. Most of the measurements were done between 8 a.m. and 5 p.m. during working days. The study area covers approximately 4 km².

For distribution analysis of these opportunistic measurements, the data were spatially aggregated within bins of 25x25m, located along the street networks and statistic of median were calculated. To assure data quality, the calculated median was based on at least 20 measurements. Maps of measurements density for each bins and for each recorded pollutant were performed (Figure 26). Similarly, maps of the distribution (median value) of the recorded pollutant concentrations within the 25m-bins were performed (Figure 27) and categorized as good, fair, moderate, poor. Most of pollutants concentration measurements fall in the moderate and poor classes in winter campaign, while in the summer campaign they fall mainly in good and fair classes.

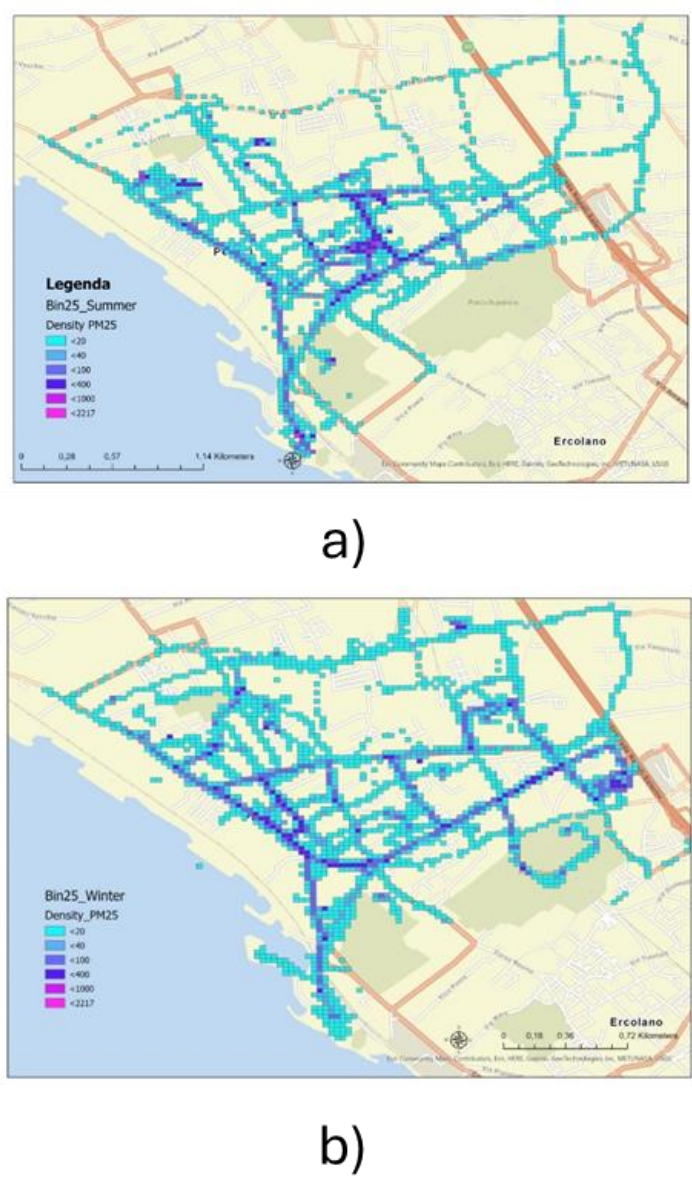
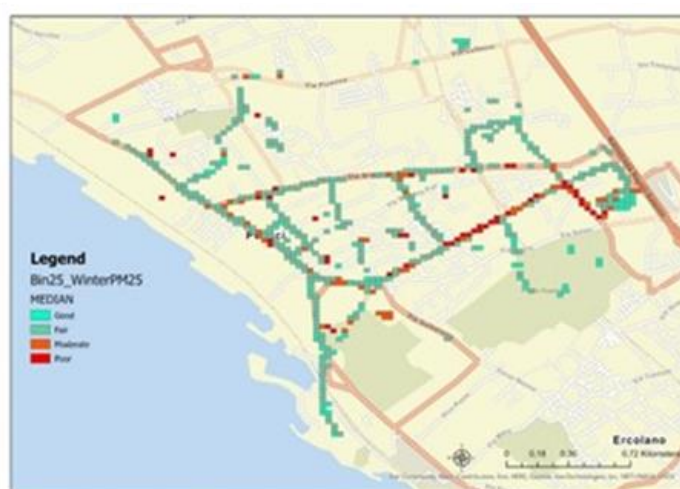


Figure 26. Maps of the PM2.5 measurements density for each 25m-bin in summer a) and winter campaigns b).



a)



b)

Figure 27. Maps of the distribution (median value) of the recorded PM_{2.5} concentrations within the 25m-bins in summer a) and winter campaigns b).

8.1. Potential Improvements for Future Geostatistical Studies

Factors that can impact on the sites eligibility for low-cost multi-sensor monitors could be considered such as the shadow effect of buildings and trees as well as the evaluation of minimum winter solar radiation value to be ensured, as further location constraints. In fact, low-cost monitors are generally equipped with a photovoltaic panel and a battery storage and installed on existing street lighting poles. Including these constraints should allow to select sites with a better sun exposure over the year. Furthermore, the use of MONICA devices at different heights as well as opportunistic and more pervasive monitoring campaigns could improve understanding of the differences between measured and modelled concentrations. Thus, a limited number of measurements and sampling errors impact the uncertainty of concentration level distributions although large spatial patterns within the city are also captured. The deployment of unstructured opportunistic mobile monitoring, coupled with repeated measurements and a meticulous approach to data processing and interpretation, is of paramount importance. This monitoring strategy must be capable of collecting a substantial amount of data, encompassing both spatial and temporal variability across the city. Until

low-cost air quality sensors are developed that can collect data of sufficient quality in an effortless manner, it is imperative to maintain motivation and follow-up with participants.

5. Conclusions and Future Directions

AirHeritage project was devised to tackle the needs of a medium sized, high density of population, city which air quality was strongly affected by anthropic effects and having no continuous regulatory grade information about pollutants concentration. The proposed technical solution was to devise, deploy and test a hierarchical, hybrid (fixed and mobile) network of low cost air quality multisensory devices mixing participatory mobile and stationary monitoring approaches so to obtain ultra high resolution spatial pollutant maps. This paper allowed to review the design choices and the outcomes deriving a set of learnt lessons which we perceive of interest of the distributed air quality sensing community at large. Feasibility, sustainability and accuracy challenges and respective proposed solutions have been discussed in different chapters devoted to Hardware choices, Software ecosystem design and implementation, Algorithmic component selection and optimization, Deployment, Operation and maintenance of the final network. All the choices have been driven by the accurate description of the project goals and operative scenario this allowed to select the appropriate network design and define the functional and operative requirements. This phase emerged undoubtedly as the most important and challenging. While accuracy it is one of the significant requirements in many scenarios, the inherent limitations of the available sensing technology must be taken into account in the aq data communication avoiding unrealistic expectance. Limitations maybe effectively tackled and remediated in the short term with adequate data driven calibration strategies, considering data gathering, access to reference stations data, relocation issues and, lastly, AI models. Despite the ad-hoc optimization of the calibration parameters the performance variance expressed by the different identically implemented measuring nodes is significant and must be taken into account during data fusion. The logistic of such ad-hoc repeated calibration solution costs may easily become prohibitive without resorting to advanced techniques. These include universal and remote calibration strategy. Both are still to be extensively tested while already commercially available. Moving resources towards continuous calibration strategy the so called golden node calibration strategy may come at hand especially when the number of measuring point is limited. Among the software functional requirement, the possibility to obtain an immediate personal exposure response has been considered the most important by users. IoT ecosystems panorama is continuously updating with new functionalities becoming available and easily deployable, among them the possibility to implement edge computing facilities should be considered paramount for LCAQMS. The use of geomatic tools has allowed to determine an optimal deployment strategy which allowed to grasp the expected modeled spatial variance with a limited number of fixed stations. We would like to stress the need for adequate preprocessing of mobile gathered data both for tackling sensors response time influence on location imputed data and for building realistic and informative pollutant concentration maps. Opportunistic measurements is desirable from the users point of view but may easily determine largely non uniform distribution in terms of location and hour of the days. Transferability of these finding is guaranteed in similar scenarios which are rather common in southern EU. In scenarios where industrial pollution become significant, a different choice of the monitored parameters will be extremely important to obtain the relevant information and to enable source apportionment. Where no fixed deployment is foreseen, establishing a mixed strategy combining opportunistic and prescheduled paths and timing can help reduce the data skewing as also emerging from [66]. We think that these insights can be of value for the community and, may help cities and companies to properly design the future monitoring network as well as carefully consider the data gathering strategy. To this aim, we hope that the sharing of the AirHeritage datalake may further help to experiment with different approaches and continue to extract value from the gathered data.

Author Contributions: Conceptualization, S.DV, G.D.; methodology: all Authors.; software, S.F., G.D., S.DV; validation, E.E., S.F., G.D.; investigation, all authors; data curation, S.F., E.E., G.D., S.DV.; writing—original draft preparation, S.F.,G.D.,S.DV, F.F., E.E., G. F.; writing—review and editing, S.DV.; visualization, S.F.,G.F., S.DV;

supervision, S.DV., G.F.; project administration, G.DF; funding acquisition, S.DV., G.DF. All authors have read and agreed to the published version of the manuscript.

Funding: This research have received partial funding by EU under UIA framework with AirHeritage Project grant no. UIA-026 and European Union's Horizon 2020 Research and Innovation programme under Project VIDIS- grant agreemen No. 952433.

Data Availability Statement: The AirHeritage datalake comprising all data generated by the AirHeritage project, as described in this present work, is made publicly available for the intended use at DOI:10.5281/zenodo.13151960 in the Zenodo repository and is indexed in OpenAire.

Acknowledgments: Authors wish to thank Dott. Gennaro Balsamo e Giuseppe Fonti for the administrative help in project management and Dr. Ezio Terzini for the unvaluable support during project proposal preparation.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. WHO, World Health Statistics 2022, Monitoring Health for the SDGs, Sustainable Development Goals World Health Organization, Geneva, Switzerland (2022)
2. M. Pascal, M. Corso, O. Chanel, C. Declercq, C. Badaloni, G. Cesaroni, S. Henschel, K. Meister, D. Haluza, P. Martin-Olmedo, S. Medina, Assessing the public health impacts of urban air pollution in 25 European cities: results of the Aphekom project, *Sci. Total Environ.*, 449 (2013).
3. Medina S., Adélaïde L., Wagner V., de Crouy Chanel P., Real E., Colette A., Couvidat F., Bessagnet B., Duron A., Host S., Hulin M., Corso M., Pascal M., Impact de pollution de l'air ambiant sur la mortalité en France métropolitaine Réduction en lien avec le confinement du printemps 2020 et nouvelles données sur le poids total pour la période 2016-2019, Saint-Maurice : Santé publique France, Paris, France (2021), p. 63
4. S. Wu, Y. Ni, H. Li, L. Pan, D. Yang, A. Baccarelli, F. Deng, Y. Chen, M. Shima, X. Guo, Short-term exposure to high ambient air pollution increases airway inflammation and respiratory symptoms in chronic obstructive pulmonary disease patients in Beijing, China, *Environ. Int.*, 94 (2016), pp. 76-82
5. O. Raaschou-Nielsen, R. Beelen, M. Wang, et al., Particulate matter air pollution components and risk for lung cancer. *Environ. Int.*, 87 (2016), pp. 66-73
6. Ute Latza, Silke Gerdes, Xaver Baur, Effects of nitrogen dioxide on human health: Systematic review of experimental and epidemiological studies conducted between 2002 and 2006, *International Journal of Hygiene and Environmental Health*, Volume 212, Issue 3, 2009, Pages 271-287, ISSN 1438-4639, <https://doi.org/10.1016/j.ijheh.2008.06.003>.
7. Kim SY, Kim E, Kim WJ. Health Effects of Ozone on Respiratory Diseases. *Tuberc Respir Dis (Seoul)*. 2020 Dec;83(Supple 1):S6-S11. doi: 10.4046/trd.2020.0154. Epub 2020 Dec 1. PMID: 33261243; PMCID: PMC7837374
8. Zhang LW, Chen X, Xue XD et al. Long-term exposure to high particulate matter pollution and cardiovascular mortality: a 12-year cohort study in four cities in northern China. *Environ Int* 62:41–47. doi:10.1016/j.envint.2013.09.012
9. Diaz-Robles L.A., Fu J.S., Reed G.D. Emission scenarios and the health risks posed by priority mobile air toxics in an urban to regional area: an application in Nashville, Tennessee. *Aerosol Air Qual Res* 13:795–803.
10. Testi, I., Wang, A., Paul, S. et al. Big mobility data reveals hyperlocal air pollution exposure disparities in the Bronx, New York. *Nat Cities* 1, 512–521 (2024). <https://doi.org/10.1038/s44284-024-00093-x>
11. N. Castell, M. Kobernus, H.-Y. Liu, P. Schneider, W. Lahoz, A. Berre, J. Noll Mobile technologies and services for environmental monitoring: the Citi-Sense-MOB approach *Urban Clim.*, 14 (2015), pp. 370-382
12. <https://www.everyaware.eu>
13. <https://www.citi-sense.eu>
14. Stojanović, D.B.; Kleut, D.; Davidović, M.; Živković, M.; Ramadani, U.; Jovanović, M.; Lazović, I.; Jovašević-Stojanović, M. Data Evaluation of a Low-Cost Sensor Network for Atmospheric Particulate Matter Monitoring in 15 Municipalities in Serbia. *Sensors* 2024, 24, 4052. <https://doi.org/10.3390/s24134052>
15. <https://www.breathelondon.org/about>
16. Nuria Castell, Franck R. Dauge, Philipp Schneider, Matthias Vogt, Uri Lerner, Barak Fishbain, David Broday, Alena Bartonova, Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?, *Environment International*, Volume 99, 2017, Pages 293-302, ISSN 0160-4120, <https://doi.org/10.1016/j.envint.2016.12.007>, (<https://www.sciencedirect.com/science/article/pii/S0160412016309989>)
17. J. Bossche, J. Theunis, B. Elen, J. Peters, D. Botteldooren, B. Baets, Opportunistic mobile air pollution monitoring: a case study with city wardens in Antwerp, *Atmos. Environ.*, 141 (2016), pp. 408-421

18. AirVisual (<https://www.indiegogo.com/projects/airvisual-node-the-world-s-smartest-air-monitor#/>)
19. Timothy M. Chukwu, Stephen Morse, Richard J. Murphy, Air quality perceptual index approach: Development and application with data from two Nigerian cities, *Environmental and Sustainability Indicators*, Volume 23, 2024, 100418, ISSN 2665-9727, <https://doi.org/10.1016/j.indic.2024.100418>.
20. UIA AirHeritage Project Description page: <https://www.uia-initiative.eu/en/operational-challenges/portici-air-heritage>. Last accessed in August, 1st, 2024.
21. Alphasense UK website: <https://www.alphasense.com>. Last accessed in August, 1st, 2024.
22. Plantower company website: <https://www.plantower.com>. Last accessed in August, 1st, 2024.
23. Teledyne API website: <https://www.teledyne-api.com/products/nitrogen-compound-instruments/t200>. Last accessed in August, 1st, 2024.
24. DIGITECO srl website: <https://www.digiteco.it>. Last accessed in August, 1st, 2024.
25. Feng, Z.; Zheng, L.; Ren, B.; Liu, D.; Huang, J.; Xue, N. Feasibility of Low-Cost Particulate Matter Sensors for Long-Term Environmental Monitoring: Field Evaluation and Calibration. *Science of The Total Environment* 2024, 945, 174089, doi:10.1016/J.SCITOTENV.2024.174089.
26. Wang, Y.; Du, Y.; Wang, J.; Li, T. Calibration of a Low-Cost PM2.5 Monitor Using a Random Forest Model. *Environ Int* 2019, 133, 105161, doi:10.1016/J.ENVINT.2019.105161.
27. Cordero, J.M.; Borge, R.; Narros, A. Using Statistical Methods to Carry out in Field Calibrations of Low Cost Air Quality Sensors. *Sens Actuators B Chem* 2018, 267, 245–254, doi:10.1016/J.SNB.2018.04.021.
28. Sá, J.P.; Chojer, H.; Branco, P.T.B.S.; Alvim-Ferraz, M.C.M.; Martins, F.G.; Sousa, S.I.V. Two Step Calibration Method for Ozone Low-Cost Sensor: Field Experiences with the UrbanSense DCUs. *J Environ Manage* 2023, 328, 116910, doi:10.1016/J.JENVMAN.2022.116910.
29. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 2016 3:1 2016, 3, 1–9, doi:10.1038/sdata.2016.18.
30. What Is REST?: REST API Tutorial Available online: <https://restfulapi.net/> (accessed on 29 June 2024).
31. Lei, Z.; Zhou, H.; Ye, S.; Hu, W.; Liu, G.P. Cost-Effective Server-Side Re-Deployment for Web-Based Online Laboratories Using NGINX Reverse Proxy. *IFAC-PapersOnLine* 2020, 53, 17204–17209, doi:10.1016/J.IFACOL.2020.12.1748.
32. Bourhis, P.; Reutter, J.L.; Vrgoč, D. JSON: Data Model and Query Languages. *Inf Syst* 2020, 89, 101478, doi:10.1016/J.IS.2019.101478.
33. How To Get Started with Node.js and Express | DigitalOcean Available online: <https://www.digitalocean.com/community/tutorials/nodejs-express-basics> (accessed on 29 June 2024).
34. Singh, J.; Chaudhary, N.K. OAuth 2.0 : Architectural Design Augmentation for Mitigation of Common Security Vulnerabilities. *Journal of Information Security and Applications* 2022, 65, 103091, doi:10.1016/J.JISA.2021.103091.
35. Auth0 Overview Available online: <https://auth0.com/docs/get-started/auth0-overview> (accessed on 29 June 2024).
36. Register Machine-to-Machine Applications Available online: <https://auth0.com/docs/get-started/auth0-overview/create-applications/machine-to-machine-apps> (accessed on 29 June 2024).
37. Change Streams & Triggers with Node.js Tutorial | MongoDB Available online: <https://www.mongodb.com/developer/languages/javascript/nodejs-change-streams-triggers/> (accessed on 29 June 2024).
38. Shahzad, F. Modern and Responsive Mobile-Enabled Web Applications. *Procedia Comput Sci* 2017, 110, 410–415, doi:10.1016/J.PROCS.2017.06.105.
39. Liu, H.; Li, Q.; Yu, D.; Gu, Y. Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms. *Applied Sciences* 2019, Vol. 9, Page 4069 2019, 9, 4069, doi:10.3390/APP9194069.
40. De Vito, S.; Esposito, E.; D'Elia, G.; Del Giudice, A.; Fattoruso, G.; Ferlito, S.; D'Auria, P.; Intini, F.; Di Francia, G.; Terzini, E. High Resolution Air Quality Monitoring with IoT Intelligent Multisensor Devices during COVID-19 Pandemic Phase 2 in Italy. 12th AEIT International Annual Conference, AEIT 2020 2020, doi:10.23919/AEIT50178.2020.9241144.
41. Building with Patterns: The Bucket Pattern | MongoDB Available online: <https://www.mongodb.com/developer/products/mongodb/bucket-pattern/> (accessed on 29 June 2024).
42. Efficient Error Monitoring: Integrating Sentry in Node.js, Express.js, and MongoDB Backend Rest API Project | Widle Studio LLP Available online: <https://medium.com/widle-studio/enhancing-error-tracking-and-monitoring-integrating-sentry-in-node-js-3f827deffca5> (accessed on 29 June 2024).
43. How To Protect an Nginx Server with Fail2Ban on Ubuntu 20.04 | Digital Ocean Available online: <https://www.digitalocean.com/community/tutorials/how-to-protect-an-nginx-server-with-fail2ban-on-ubuntu-20-04> (accessed on 29 June 2024).

44. Is Protobuf Much Faster than Json Even in Simple Web Server Response Requests? | by Kuentra Official | Medium Available online: <https://medium.com/@kn2414e/is-protocol-buffers-protobuf-really-lighter-and-faster-compared-to-json-681c6bee5d93> (accessed on 29 June 2024).
45. Masdani, M. V.; Darlis, D. A Comprehensive Study on MQTT as a Low Power Protocol for Internet of Things Application. *IOP Conf Ser Mater Sci Eng* 2018, 434, doi:10.1088/1757-899X/434/1/012274.
46. Time Series Collection Limitations - MongoDB Manual v7.0 Available online: <https://www.mongodb.com/docs/manual/core/timeseries/timeseries-limitations/> (accessed on 29 June 2024).
47. Improving Scalability of Cloud-Based IoT Application with Microservice Architecture Using Massively Distributed Kubernetes Cluster | Request PDF Available online: https://www.researchgate.net/publication/346443549_Improving_Scalability_of_Cloud-based_IoT_Application_with_Microservice_Architecture_using_Massively_Distributed_Kubernetes_Cluster (accessed on 29 June 2024).
48. Rani, F.; Wang, X.; Mutlu, I.; Urbas, L. A Centralized MongoDB-Based Repository Design for IIoT Data: The EcoKI Project*. *IFAC-PapersOnLine* 2023, 56, 3184–3189, doi:10.1016/j.ifacol.2023.10.1454.
49. Schneider, P.; Vogt, M.; Haugen, R.; Hassani, A.; Castell, N.; Dauge, F.R.; Bartonova, A. Deployment and Evaluation of a Network of Open Low-Cost Air Quality Sensor Systems. *Atmosphere* 2023, 14, 540. <https://doi.org/10.3390/atmos14030540>
50. Malings, C.; Tanzer, R.; Haurlyuk, A.; Kumar, S.P.; Zimmerman, N.; Kara, L.B.; Presto, A.A.; Subramanian, R. Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. *Atmos. Meas. Tech.* 2019, 12, 903–920.
51. S. De Vito, G. D'Elia and G. Di Francia, "Global calibration models match ad-hoc calibrations field performances in low cost particulate matter sensors," 2022 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), Aveiro, Portugal, 2022, pp. 1-4, doi: 10.1109/ISOEN54820.2022.9789669.
52. Smith, K.R.; Edwards, P.M.; Ivatt, P.D.; Lee, J.D.; Squires, F.; Dai, C.; Peltier, R.E.; Evans, M.J.; Sun, Y.; Lewis, A.C. An Improved Low-Power Measurement of Ambient NO₂ and O₃ Combining Electrochemical Sensor Clusters and Machine Learning. *Atmos. Meas. Tech.* 2019, 12, 1325–1336.
53. Timothy M. Swager, and Katherine A. Mirica, Introduction: Chemical Sensors, *Chem. Rev.* 2019, 119, 1, 1–2 Publication Date: January 9, 2019 <https://doi.org/10.1021/acs.chemrev.8b00764>
54. G. D'Elia, S. De Vito, P. Sommella e M. Ferro, Influence of Concept Drift on Metrological Performance of Low-Cost NO₂ Sensors, *IEEE Transactions on Instrumentation and Measurement*, pp. 1-11, 2022
55. G. Casey e M. P. Hannigan, «Testing the performance of field calibration techniques for low-cost gas sensors in new deployment locations: across a county line and across Colorado,» *Atmospheric Meas. Tech.*, p. 6351–6378, 2018.
56. S. Vikram, A. Collier-Oxandale, M. H. Ostertag, M. Menarini, C. C. S., T. R. Dasgupta, M. Hannigan e W. G. Griswold, «Evaluating and improving the reliability of gas-phase sensor system calibrations across new locations for ambient measurements and personal exposure monitoring,» *Atmospheric Meas. Tech.*, p. 4211–4239, 2019
57. De Vito, S.; Esposito, E.; Massera, E.; Formisano, F.; Fattoruso, G.; Ferlito, S.; Del Giudice, A.; D'Elia, G.; Salvato, M.; Polichetti, T.; et al. Crowdsensing IoT Architecture for Pervasive Air Quality and Exposome Monitoring: Design, Development, Calibration, and Long-Term Validation. *Sensors* 2021, 21, 5219. <https://doi.org/10.3390/s21155219>
58. S. De Vito et al., A Global Multiunit Calibration as a Method for Large-Scale IoT Particulate Matter Monitoring Systems Deployments," in *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-16, 2024, Art no. 2501916, doi: 10.1109/TIM.2023.3331428
59. Zimmerman, N.; Presto, A.A.; Kumar, S.P.N.; Gu, J.; Haurlyuk, A.; Robinson, E.S.; Robinson, A.L.; Subramanian, R. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmos. Meas. Tech.* 2018, 11, 291–313
60. Marie Ramel-Delobel, Cosimo Peruzzi, Thomas Coudon, Saverio De Vito, Grazia Fattoruso, Delphine Praud, Béatrice Fervers, Pietro Salizzoni, Exposure to airborne particulate matter during commuting using portable sensors: Effects of transport modes in a French metropolis study case, *Journal of Environmental Management*, Volume 365, 2024, 121400, ISSN 0301-4797, <https://doi.org/10.1016/j.jenvman.2024.121400>
61. G. Miskell, K. Alberti, B. Feenstra, G. S. Henshaw, V. Papapostolou, H. Patel, A. Polidori, J. A. Salmond, L. Weissert e David E. Williams, «Reliable data from low cost ozone sensors in a hierarchical network,» *Atmospheric Environment*, n. 214, 2019
62. S. D. Vito, G. D. Elia, S. Ferlito, E. Esposito, G. Piantadosi and G. D. Francia, "Remote Calibration strategies for Low Cost Air Quality Multisensors: a performance comparison," 2024 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), Grapevine, TX, USA, 2024, pp. 1-4, doi: 10.1109/ISOEN61239.2024.10556107

63. Fattoruso, G.; Nocerino, M.; Toscano, D.; Pariota, L.; Sorrentino, G.; Manna, V.; De Vito, S.; Carteni, A.; Fabbricino, M.; Di Francia, G. Site Suitability Analysis for Low Cost Sensor Networks for Urban Spatially Dense Air Pollution Monitoring. *Atmosphere* 2020, 11, 1215. <https://doi.org/10.3390/atmos11111215>)
64. Soulhac, L.; Salizzoni, P.; Cierco, F.X.; Perkins, R. The model SIRANE for atmospheric urban pollutant dispersion; PART I, presentation of the model. *Atmos. Environ.* 2011, 45, 7379–7395.
65. Fattoruso, G.; Toscano, D.; Cornelio, A.; De Vito, S.; Murena, F.; Fabbricino, M.; Di Francia, G. Using Mobile Monitoring and Atmospheric Dispersion Modeling for Capturing High Spatial Air Pollutant Variability in Cities. *Atmosphere* 2022, 13, 1933. <https://doi.org/10.3390/atmos13111933>
66. Jonathan Gillespie, Nicola Masey, Mathew R. Heal, Scott Hamilton, Iain J. Beverland, Estimation of spatial patterns of urban air pollution over a 4-week period from repeated 5-min measurements, *Atmospheric Environment*, Volume 150, 2017, Pages 295-302, ISSN 1352-2310, <https://doi.org/10.1016/j.atmosenv.2016.11.035>. (<https://www.sciencedirect.com/science/article/pii/S1352231016309086>)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.