Doctoral Dissertation
Doctoral Program in Computer Engineering and Control (35^{th}cycle)

IoT for Urban Sustainability in Smart Cities

By

Gustavo Adolfo Ramírez Espinosa

******

Supervisor(s):
Prof. Bartolomeo Montrucchio

Doctoral Examination Committee:
Prof. Claudio Zunino, Referee, Istituto di Elettronica e Ingegneria dell'Informazione e delle Telecomunicazioni (IEIIT), (IT)
Prof. Diego Méndez Chavez, Referee, Pontificia Universidad Javeriana, (COL)
Prof. Claudio Fornaro, Università Telematica Internazionale Uninettuno (IT)
Prof. Andrea Sanna, Politecnico di Torino (IT)
Prof. Renato Ferrero, Politecnico di Torino (IT)

Politecnico di Torino
2023

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Gustavo Adolfo Ramírez Espinosa
2023

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

I would like to dedicate this thesis to my loving parents, my Love, my dogs and my friends. All of you have provided me with the support to reach this goal.

Acknowledgements

I would like to express my gratitude to the many people who were involved in my educational journey. This has been an incredible journey, and words alone cannot adequately convey my gratitude.

First and foremost, I want to thank my parents for enabling me to fulfill a dream and for filling them with pride. Their sacrifices and the tools they provided me with have allowed me to reach this milestone.

I extend my heartfelt thanks to my supervisor for giving me the opportunity to pursue this goal, as well as for their guidance and invaluable advice.

I am grateful to Edoardo, who has been more than just a good friend but also a mentor, guiding me through this unfamiliar process.

My thanks go to my boss, Diego, for believing in me and giving me the opportunity to have a tenured position.

I want to express my appreciation to my entire research team for their contributions to the implementation and our daily work toward a common objective.

I am thankful to my Turin family, and friends, who made my life easier and helped me adapt to this new country.

To my Love, who has always seen the best in me, showered me with love, and patiently waited for me – I will always be grateful. Thanks to my dogs, to show me the peace and happiness.

I would also like to acknowledge Politecnico di Torino, Pontificia Universidad Javeriana, and all the companies that supported my research.

I also extend my gratitude to my Italian friends, and all the Italian people. I will never forget your kindness and assistance.

Abstract

In the highly interconnected world of today, new technologies that allow for the digitalization of the physical world are brought to the forefront of discussion. The Internet of Things (IoT) and mature Information and Communication Technologies (ICT) have driven various digital and industrial transformation processes (Industry 4.0). Meanwhile, humanity faces emerging challenges that threaten its well-being and the environment, with recent issues like pandemics and climate emergencies capturing the attention of governments and scientific communities.

In this context, IoT and ICT gain significance as an emerging solution for diagnosing and making decisions regarding problems that threaten city sustainability. The implementation of these technologies as robust infrastructure has given rise to the concept of smart cities or intelligent societies. However, the real value does not lie in the technologies themselves, but in the ***data*** they generate. Deploying IoT solutions is not trivial, as designing both an IoT device and a system that collects valuable information involves careful consideration from multiple perspectives.

In IoT, the tasks of device deployment and data acquisition, storage, and analysis each present their own challenges. The state-of-the-art is rich in solutions seeking to contribute to each of these issues. This thesis report aims to undertake a comprehensive analysis covering the main problems that a solution could encounter. From studying the scenario and technical feasibility of the task at hand; designing a flexible device that allows adaptability to various types of measurements; determining the correct frequency of data sampling and efficient use of device resources; the type of communication and network deployment needed to define the solution's coverage; and ensuring the quality, credibility, and coherence of data generated by devices, as well as protecting it from potential alterations.

The analysis conducted in this thesis has also opened up new questions about tangential scenarios related to certain dynamics, such as the use of architectures that

merge IoT with emerging technologies like blockchain. This is relevant not only to the subject of interest but also as an alternative for issues like identity theft and citizen impersonation. Moreover, considering the constraints and challenges faced during the pandemic in education, an exploratory scenario is proposed where IoT could serve as a supportive tool for teaching. This would facilitate the quantification of parameters such as student attention.

Summarizing, the main contributions of this thesis are the following:

- *IoT system for the quantification of human well-being parameters in indoor and outdoor environments of a building.*
  The subject matter of this research focuses on outdoor air quality measurement, indoor building comfort, and attention levels. Different IoT solutions are proposed for each application and evaluated as proofs of concept for the initial quantification of these phenomena.

- *Optimal sampling in terms of energy and data quantity for particulate matter pollution variables in low-cost sensors*.
  This work examines the frequency dynamics of particulate matter pollution in urban environments, defining optimal sampling frequencies and studying the effects of applying duty-cycling techniques to reduce sensor power consumption. The impact of these techniques on sensor precision is also assessed.

- *An architecture for hybrid deployments using various wireless technologies.*
  Various deployment topologies and different wireless transmission technologies for transmitting particulate matter pollution data are explored. Each technology's unique constraints are considered with the aim of enhancing spatio-temporal resolution of measurements within a city.

- *A process for generating scientifically valid data for low-cost sensors.*
  This topic investigates numerous algorithms encapsulated within a framework designed to achieve precise measurements, comparable with reference measurement equipment, using low-cost sensors known for their noisy behavior and high failure rates.

- *Integration of IoT and blockchain for data integrity protection.*
  The study further analyzes two use cases of Distributed Ledger Technology (DLT) as a solution for the protection of both personal and public data, as well as city measurement data. Hybrid models are utilized to minimize resources.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The way we interact with technology and the world around us has been revolutionized by the Internet of Things (IoT) and Information and Communication Technology (ICT). A new era of connectivity, intelligence, and efficiency has been brought about by the convergence of these two fields. Valuable information is generated by these devices, which play a significant role in our daily lives and can be utilized for decision-making and lifestyle adaptation. The IoT is considered a crucial enabler of the so-called Digital Transformation or Industry 4.0, which encompasses the transformation of industries across various sectors, including public administrations and their entities [1].

To gain an understanding of the value that the IoT will contribute in this digital transformation context, studies such as [2] estimated that by 2025, the global impact value of the Internet of Things will range between 3.9 trillion and 11.1 trillion dollars across nine sectors, including transportation (land and maritime), healthcare, commerce, manufacturing, and cities.

The Internet of Things refers to a vast network of interconnected devices, objects, and systems that have the ability to collect, exchange, and analyze data. These devices, often embedded with sensors and actuators, can communicate with each other and with humans through the internet. This interconnectedness allows for seamless integration of the physical and digital worlds[3]. On the other hand, Information and Communication Technology encompasses the technologies and tools used to manage and communicate information. It encompasses a broad spectrum

of technologies, including computers, networks, software, and telecommunications, that enable the collection, storage, processing, and transmission of data.

Through the power of connectivity, IoT devices can gather real-time data, monitor and control physical systems, optimize processes, and enhance decision-making. Meanwhile, ICT plays a crucial role in providing enabling technologies that support the infrastructure, communication protocols, and data management systems necessary for the functioning of IoT. ICT provides the backbone for data transmission, storage, and analysis, enabling the seamless flow of information between devices and users.

The increasing adoption and proliferation of IoT platforms can be attributed to three distinct factors:

- *Internet Accesss*: The proliferation of heterogeneous communication networks, along with their affordability in terms of costs and coverage, is particularly notable in urban sectors. This phenomenon has contributed to the widespread adoption and deployment of IoT technologies in urban environments.

  In recent years, advancements in communication technologies have led to the availability of diverse network options, such as cellular networks, and LPWAN (Low-Power Wide Area Networks), Wireless and Wire Area Network, and Wireless Personal Networks. These networks offer varying levels of data transmission speeds, coverage range, and power consumption, catering to different IoT application requirements. Furthermore, the reduced costs associated with deploying communication infrastructure have made it more accessible for businesses, organizations, and even individuals to implement IoT solutions [4]. This has spurred the growth of IoT deployments, especially in urban areas, where the infrastructure is already well-established.

  The combination of affordable and diverse communication networks, coupled with the high population density and demand for innovative services in urban sectors, has created a fertile ground for the expansion of IoT applications.

- *Technological maturity and affordability*: Advancements in technology have led to significant reductions in costs, size, and energy consumption of IoT components[5]. These developments have played a crucial role in driving the widespread adoption of IoT solutions. Reduced costs have made it economically feasible to integrate IoT capabilities into various devices, while miniaturization has enabled IoT integration into smaller devices. Improved

energy efficiency has extended battery life and expanded deployment possibilities.

- *Advancements in data storage and processing technologies*: They have also played a pivotal role in enabling the widespread utilization of IoT-generated data. The increased capabilities and reduced costs of storage solutions, coupled with powerful processors and advanced algorithms, have facilitated the efficient handling and analysis of massive datasets. These developments have not only empowered industries to leverage IoT-generated data for valuable insights but have also made these technologies more accessible and scalable across various domains.

1.1 The IoT Architecture

In [6], an IoT architecture divided into four subsystems is presented, which are defined as layers and showed in Figure1.1.

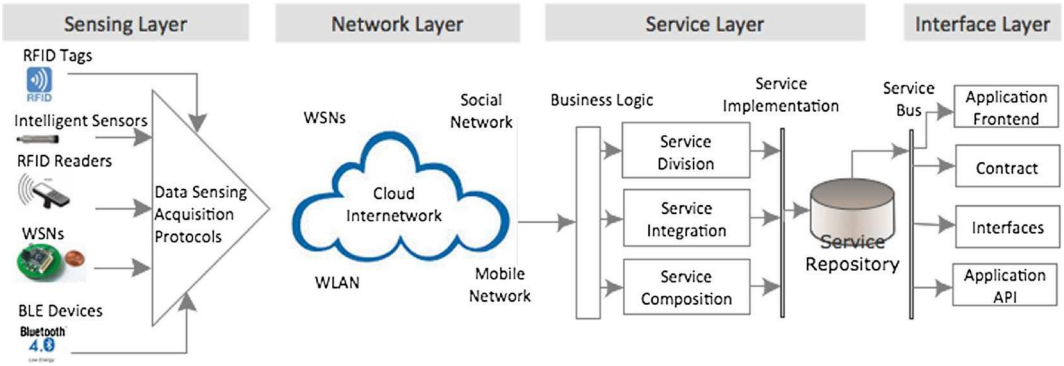


Fig. 1.1 IoT general architecture [6]

The first layer is the sensor layer, responsible for gathering information from the environment through hardware devices such as sensor nodes. This layer enables the IoT system to sense and collect data from various sources.

The second layer is the network layer, which establishes the connectivity and communication among the IoT devices. It facilitates the seamless exchange of data and enables the things in the network to be aware of their surroundings. The network layer plays a crucial role in enabling intelligent event management and processing within the IoT system.

The third layer is the service layer, which relies on middleware technology to provide services and applications in the IoT environment. This layer involves activities related to service specifications, discovery, composition, and management. It acts as a platform for executing IoT services and ensures the interoperability and reusability of hardware and software components.

The fourth layer is the interface layer, which addresses the compatibility and interaction issues among heterogeneous IoT devices. It provides an effective mechanism for simplifying the management and interconnection of things. Interface profiles are used to describe the specifications between applications and services, enabling seamless interactions among different devices.

By dividing the IoT architecture into these four layers, the proposed framework in [6] aims to ensure the interconnection, extensibility, scalability, and interoperability of IoT devices and services. It provides a comprehensive structure that bridges the gap between the physical and virtual worlds, enabling efficient and dynamic interactions among IoT components.

1.2 IoT and Urban Environments

The integration of IoT technologies in urban environments is generating significant interest among researchers and practitioners. This development is expected to have a profound impact on urban sustainability, as it holds the potential to enhance efficiency, promote sustainability, and improve the overall quality of life in cities.

In the pursuit of Smart Cities, where ICT solutions are utilized for effective public management, the implementation of IoT in urban contexts becomes particularly relevant[7]. The overarching objective remains the optimization of public resources, the provision of high-quality services to citizens, and the reduction of operational costs for public administrations. Achieving these goals can be facilitated through the establishment of urban IoT infrastructures that offer unified, accessible, and cost-effective access to a wide range of public services. Such an approach fosters collaboration, synergies, and transparency, benefiting citizens and the urban environment.

The deployment of an urban IoT has the potential to revolutionize the management and optimization of traditional public services in cities, including transportation

and parking management, efficient lighting systems, public area surveillance and maintenance, cultural heritage preservation, waste management, hospital sanitation, and school facilities[8]. Moreover, the wealth of diverse data collected through an extensive urban IoT network can be leveraged to increase transparency, engage citizens, raise awareness about the city's condition, encourage citizen participation in public administration, and stimulate the development of innovative services built upon the IoT infrastructure[9].

Several studies have been conducted under the concept of smart cities using IoT, and some of these projects, serving as case studies, are summarized in [8]. Here, there is presented a list of relevant case studies:

- **Waste management**: Waste management poses significant challenges in modern cities, including high costs and the issue of landfill storage. An example of this is the implementation of intelligent waste containers that can detect their fill levels, enabling the optimization of collection routes. By leveraging IoT technology, these containers are connected to a central control center where data is processed by optimization software to determine the most efficient management of the waste collection fleet[10].

- **Noise pollution**: It is a significant issue in urban environments, comparable to the impact of carbon oxide (CO) on air quality. To address this concern, city authorities have implemented specific regulations to reduce noise levels during certain hours in city centers. The integration of urban IoT technology can provide a noise monitoring service, measuring and tracking noise levels at different locations and times [11]. This service not only enables the creation of a space-time map of noise pollution but also supports public security efforts by employing sound detection algorithms capable of identifying specific sounds like glass breaking or altercations.

- **Traffic congestion**: Another valuable service enabled by urban IoT in a Smart City context is the monitoring of traffic congestion. While camera-based traffic monitoring systems are already in use in many cities, the widespread deployment of low-power communication networks can offer a more extensive and detailed source of information. Traffic monitoring can leverage the sensing capabilities and GPS technology present in modern vehicles [15], along with a combination of air quality and acoustic sensors installed along specific roads.

Authorities can utilize it to manage traffic flow effectively, deploying personnel where needed, while citizens can plan their routes in advance for commuting or shopping in the city center.

- **Comfort and well-being in indoor environments**: The monitoring of energy consumption and environmental conditions in public buildings, such as schools, administration offices, and museums, using IoT technologies, is another significant application. This involves the use of various sensors and actuators to regulate lighting, temperature, and humidity levels. By managing these factors, it becomes feasible to improve the comfort of occupants, potentially boosting productivity and reducing heating/cooling expenses [12]. This application showcases how IoT can contribute to optimizing building operations and enhancing the well-being of individuals in public spaces.

- **Air quality**: The European Union has officially adopted different directives, which set ambitious goals for reducing climate change impact. These targets include a reduction in greenhouse gas emissions, a decrease in energy consumption through improved efficiency, and a increase in the use of renewable energy sources. In line with these objectives, the implementation of an urban IoT infrastructure can play a crucial role in monitoring air quality in densely populated areas, parks, and fitness trails [13]. This technology can also facilitate communication between health applications and the IoT infrastructure, enabling individuals to choose the healthiest routes for outdoor activities. Achieving this service requires the deployment of air quality and pollution sensors throughout the city, with the collected data being made accessible to the public.

1.3 Open Challenges

The Internet of Things (IoT) is still an emerging paradigm, and it faces several technological challenges that need to be addressed through research. Various studies have classified these challenges from the perspective of the paradigm itself [6, 3], architectural considerations [14], or application domains such as smart cities [15]. Some prominent challenges include:

- Interoperability and standardization

- Scalability
- Flexibility
- Energy efficiency
- Mobility management
- Security
- Privacy protection
- Information reliability
- Low-cost deployment (Devices and communications)
- Intelligent and relevant analytics

The aforementioned challenges outlined are, in broad terms, some of the challenges that generate interest in future research directions. In each of these points, there is great potential as the application of IoT in different contexts demands diverse needs that still require methodological and technological attention in research.

These challenges represent opportunities for further investigation and development in the field of IoT. The diverse nature of IoT applications necessitates a deep understanding of specific contexts and the development of tailored methodologies and technologies. By addressing these challenges, researchers can contribute to the advancement of IoT and pave the way for its successful implementation in various domains

1.4 Research Questions

The present thesis addresses some of these aspects that are often seen in a general manner when proposing a solution or deployment of IoT in urban contexts. In particular, these aspects can be prioritized according to the architecture outlined in Section 1.1 When analyzing the flow of information, it is identified that one of the critical points in data acquisition is the sensing layer.

In this part, various sensors are involved, which are responsible for digitizing different physical variables. However, the ability to deploy at a large scale at a low

cost while maintaining the required spatiotemporal resolution always involves a trade-off among factors such as precision, reliability, energy consumption, resolution, and cost. Finding an optimal point for each factor constitutes a significant challenge in itself, which is highly dependent on the application.

Poor optimization in this layer can hinder the optimal functioning of the rest of the architecture, as it can create pressure on resource demand in the subsequent layers. For example, massive transmission of data that lacks meaningful information (as described by Shannon's information theory) and unnecessary or unproductive processing in the service layer. As a result, the data delivered at the interface layer may fail to generate value.

Efficient optimization of the sensing layer is crucial to ensure the overall effectiveness and efficiency of the IoT architecture. By carefully selecting appropriate sensors, designing efficient data acquisition strategies, and considering the specific requirements of the application. This, in turn, will enable the delivery of valuable and meaningful data at the interface layer, leading to enhanced decision-making processes and improved performance of IoT systems in urban contexts.

Similarly, in the network and service layers, the lack of optimization can affect the delivery of information or fail to properly detect patterns that generate a detailed understanding of what is happening in the physical world.

In the network layer, the efficient transmission and routing of data play a critical role in ensuring timely and reliable communication between IoT devices and the backend infrastructure. Inadequate network optimization can result in delays, packet losses, or bottlenecks that hinder the seamless flow of information. This can lead to incomplete or inaccurate data being received.

In the service layer, optimization is essential for processing and analyzing the collected data effectively. This involves leveraging appropriate algorithms, machine learning techniques, and computational resources to extract valuable insights and identify meaningful patterns from the data. Insufficient optimization in this layer can result in inaccurate analysis, missed opportunities for real-time decision-making, and a lack of detailed understanding of the physical world.

To address these challenges the present thesis want to answer the following questions to enhance the overall performance of IoT systems:

- *Could a set of low-cost sensors provide information that can determine human comfort levels or understand specific factors in the complexity of an urban environment?*
- *Do the sensors sample at the correct frequency to optimally digitize the PM behavior??*
- *Is it possible to reduce sensor power consumption without degrading accuracy?Is it possible to reduce the power consumption of the sensors? What are the implications for their precision?*
- *What communication protocol is most suitable for transmitting data in urban environments?*
- *Are the data generated by the sensors reliable and scientifically valuable?*
- *Can citizens trust in the integrity and authenticity of the data?*
- *Can IoT be utilized to explore other aspects of human well-being, specifically those related to mental focus?*

These questions are addressed throughout this document within the context of some of the challenges presented in section 1.3, where urban sustainability serves as a general case study. In the following subsections, an overview of all chapters included in this thesis will be provided.

1.4.1 IoT Environments' Domain

This chapter explores the integration of IoT technologies in addressing environmental issues and promoting well-being in urban environments. Additionally, this chapter takes a first approximation to examine the potential of low-cost sensors in providing valuable information and insights through two case studies. The research presented in this chapter contributes to two key areas: comfort and well-being in indoor environments, and air quality in outdoor environments.

Considering the significance emphasized by the United Nations' Sustainable Development Goals (SDGs), particularly Goal 11 aimed at achieving inclusive, safe, resilient, and sustainable cities and human settlements, this study explores the

question of whether a set of low-cost sensors can provide the necessary information to determine levels of human comfort or explain healthy environments in urban contexts. Then, case studies are presented in the chapter to highlight the importance of comfort and well-being in indoor environments, as well as outdoor air quality, specifically in the monitoring of ultra-fine particles. These case studies investigate the potential of low-cost sensors in providing insights and understanding the complex factors at play. The effectiveness of low-cost sensors in enhancing the well-being and quality of life for city dwellers is assessed by examining human comfort levels and specific aspects of urban environments.

In the context of comfort and well-being in indoor environments, the research conducted in this field is being introduced, specifically on the development and characterization of a low-cost multi-sensor device called PROMET&O. The development and characterization of PROMET&O, designed to monitor Indoor Environmental Quality (IEQ) parameters and provide valuable insights into human comfort levels, are emphasized. The aim of this research is to assess whether a better understanding of the complexities within urban environments and an improvement in indoor comfort conditions can be effectively contributed by low-cost sensors.

Moving on to air quality, the implementation of an urban IoT monitoring campaign for monitoring ultra-fine particles in a densely populated residential area is being explored. By deploying low-cost sensors throughout the city, data and insights into particulate levels and their interaction with the urban fabric. The capabilities and limitations of low-cost sensors in providing valuable information for creating healthier and more sustainable cities are investigated through the research

By addressing the question of whether a set of low-cost sensors can provide information to determine human comfort levels and understand specific factors in the complexity of urban environments, this chapter aims to shed light on the potential of IoT applications. The integration of IoT technologies, guided by the SDGs, offers promising opportunities to leverage low-cost sensors in addressing environmental challenges and enhancing the well-being of urban populations. Through the examination of case studies and research contributions, we strive to assess the effectiveness and applicability of low-cost sensors in promoting sustainable and healthy smart cities.

1.4.2 Data Collection

Now, with a better understanding of the application environment, the monitoring of air quality is taken as the main case study, focusing specifically on the particulate matter pollutant. This pollutant is generated through the combustion of fossil fuels, industrial production, and various combustion processes. These ultrafine particles are classified based on their diameter.

- PM_{10} refers to particles with a diameter smaller than $10\mu m$ and larger than $2.5\mu m$.
- $PM_{2.5}$ represents particles with a diameter smaller than $2.5\mu m$ and larger than $1.0\mu m$.
- $PM_{1.0}$ denotes particles with a diameter smaller than $1.0\mu m$.

These micro-particles have a significant impact on the human respiratory system, causing various respiratory diseases and increasing the risk of lung cancer. Urban environments are major sources of particulate matter due to the concentration of construction activities, combustion-based transportation, and industries. Consequently, their influence on human well-being has gained significant attention in recent years, leading to an increased interest in their monitoring [16]. However, monitoring particulate matter is not a trivial task, as determining its concentration in the air requires complex processes [13]. Under European directives (2008/50/EC), two techniques are approved for the official measurement of these parameters: β-attenuation monitoring and gravimetric detection. Both techniques employ filters to trap particles, and the concentration of the pollutant is determined based on the attenuation of β radiation or weight changes (gravimetric).

Both techniques are costly, affecting spatial resolution, and provide measurements with low temporal resolution. As an alternative, low-cost sensors based on light scattering have emerged as more affordable options, although their accuracy is still limited. However, they can contribute to improving spatial and temporal resolutions. These sensors can generate data at speeds of one measurement per second, but they present challenges in the context of IoT. Their measurements require a constant airflow, which is achieved through the use of a fan or a resistor to manipulate the air temperature. This makes them energy-intensive. Moreover, it is unclear whether such a high sampling rate truly provides useful information.

Given the aforementioned challenges, this chapter aims to explore the efficient utilization of these types of sensors by addressing two key questions: First, whether the sensors operate at the correct frequency by analyzing the behavior of the physical variable and its frequency components. And second, whether it is possible to reduce energy consumption by applying techniques such as duty cycling to achieve reduced sensor power consumption, while also analyzing the implications for measurement accuracy.

Overall, this chapter contributes to the advancement of IoT-based air monitoring systems by addressing the challenges associated with low-cost sensor acquisition. The insights and findings provide valuable guidance for optimizing the performance, energy efficiency, and data quality of these systems, enabling more effective air pollution monitoring and mitigation strategies in smart city environments.

1.4.3 Networks and Data Transmission

The emergence of IoT has led to a significant change in the way devices exchange information. Wireless connectivity plays a vital role in the success of IoT, facilitating smooth communication between a wide range of devices, sensors, and systems. It is essential to comprehend the importance of wireless connectivity in IoT to fully exploit the transformative capabilities of this technology.

However, there are currently several technologies that provide wireless communication, each with its own strengths and weaknesses. In the field of IoT, there are various options where the type of scenario in which they are deployed takes precedence, as each IoT deployment entails different parameters to consider, defining requirements that must be met by the technology or protocol that optimizes communication [17, 18]..

In general, the key requirements for wireless communication in an IoT deployment can be summarized as follows: coverage, scalability, reliability, power consumption, data throughput, interoperability, and security.

Scalability, as the number of connected devices is projected to reach billions in the near future, it is imperative that the wireless infrastructure is capable of accommodating the exponential growth of these endpoints. Whether it involves a smart home, industrial facility, or smart city, the wireless network must be engineered

to effectively manage the substantial quantity of devices and their data transmission demands, all while maintaining optimal performance levels.

Coverage pertains to the extent of wireless connectivity within a geographical region. In the context of IoT, applications can encompass expansive territories, ranging from urban settings to rural or isolated areas. Wireless technologies utilized in IoT must offer broad coverage to ensure connectivity is accessible across these diverse locations.

Power consumption is a crucial factor to consider in wireless connectivity for IoT. Numerous IoT devices operate on battery power and are often deployed in remote or hard-to-reach locations. In order to maximize the operational lifespan of these devices and minimize maintenance efforts, wireless technologies must prioritize power optimization. Employing efficient power management techniques and utilizing low-power communication protocols are imperative for enabling long-lasting and energy-efficient IoT deployments.

The **reliability** of wireless networks is paramount in IoT applications that encompass mission-critical functions such as healthcare monitoring, industrial automation, or transportation systems. To enable uninterrupted data transfer and real-time response, wireless networks must provide consistent and dependable connectivity. This requires the implementation of robust wireless protocols and reliable network architectures, which are crucial in facilitating seamless and reliable communication among devices, even in challenging environments or areas with significant interference.

Data throughput is an essential factor to consider in wireless communications for IoT. While IoT devices commonly transmit small data payloads, specific applications require higher data rates or deal with substantial data volumes. It is imperative for IoT deployments to carefully select wireless technologies capable of efficiently handling low-rate, intermittent data transmission or high-rate continuous data flow. This selection process ensures that the chosen wireless technology can accommodate the diverse range of IoT applications and effectively address their unique data requirements.

Interoperability is another critical requirement for wireless connectivity in the context of IoT. The vast ecosystem of IoT devices and platforms demands wireless technologies that can seamlessly integrate and interoperate across diverse devices, protocols, and networks. Standardized wireless protocols and interoperable

frameworks play a pivotal role in enabling communication and collaboration among devices from various manufacturers, fostering an open and flexible IoT ecosystem.

Security cannot be overstated in IoT, given the potential sensitivity of the transmitted data. IoT systems could handle vast quantities of personal, industrial, and critical infrastructure data, making them vulnerable to malicious attacks. To protect IoT communications from unauthorized access, robust encryption, authentication, and data integrity mechanisms are crucial. These security measures ensure the privacy and integrity of data transmitted over wireless networks, safeguarding against potential breaches and unauthorized manipulation.

By addressing the previous requirements, wireless technologies can provide comprehensive connectivity solutions for IoT deployments, ensuring seamless operation and real-time decision-making capabilities.

These requirements serve as the basis for constructing resilient and effective wireless networks capable of supporting the varied and continuously expanding IoT environment. By comprehending and addressing these requirements, businesses, industries, and individuals can unleash the complete potential of IoT, guaranteeing smooth operations and enabling real-time decision-making capabilities.

In this chapter, a case study focused on urban particulate matter monitoring is presented to address two key questions: which communication protocol is most suitable for transmitting data in urban environments, and is it possible to optimize data transmission? The chapter emphasizes the challenges associated with data transmission and explores a variety of data transmission schemes facilitated by diverse wireless transmission technologies.

These schemes utilize the monitoring station discussed in section 3.2. The design for the monitoring station, presented in this chapter, is characterized by its connectivity flexibility, capable of functioning under three distinct network topologies: Fixed Sensor Networks, Participatory Sensor Networks, and Mobile Sensor Networks.

Furthermore, strategies for optimizing data transmission in low throughput scenarios, especially those involving BLE and LoRa technologies, are introduced in this chapter.

1.4.4 Data Quality

Low-cost sensors still present numerous challenges in terms of precision, coherence, and resilience. The operating principles of certain sensors make them more susceptible to aging or environmental factors that degrade their performance.

Specifically, the particulate matter sensors primarily used in the research presented in this document operate based on the light scattering principle (LSP) caused by ultrafine particles crossing a light beam. However, the model used for calculating particulate concentration is often based on specific types of particles, such as cigarette smoke. This calibration may not accurately measure other types of particles, so using these sensors outdoors requires recalibration [19].

Furthermore, these light-scattering-based sensors require a constant airflow, leading many to incorporate resistors to generate air convection or employ a small fan. Both methods require significant energy, and in the case of the fan, introduce another potential point of failure. Despite this, sensors with fans tend to perform better as they ensure more stable airflow and a more rapid response time.

Another performance-affecting factor is humidity, which can alter particle density or cause internal condensation that disrupts the light scattering sensor reading [13]. The sensor's deployment time is also a consideration, as outdoor exposure to dust can reduce its lifespan and allow particles like insects, seeds, or dust to clog the sensor.

These issues represent significant challenges in deploying IoT solutions based on these types of sensors, given their high failure probability and potential inconsistency between sensors. Therefore, this chapter explores how to ensure the data collected by these sensors is both reliable and scientifically valid.

We adopt the concept of redundant sensors within a single IoT device, which is calibrated against a professional monitoring station. Following this calibration, several calibrated devices are deployed for a year to assess the quality of their data compared to the reference station. To discard anomalous values and failures presented by some sensors, we propose a framework that filters out failed sensors and eliminates anomalous data with the aim of reducing error and enhancing the coherence of the sensor data.

1.4.5 Data Integrity

Once the data has been processed in such a way that the air monitoring stations provide values close to the reference values, the issue of maintaining data immutability arises and how to make them openly available to citizens and interested third parties. As the data is publicly available and reflects some of the processes taking place in a city that contribute to the welfare of the inhabitants, ensuring its integrity becomes crucial.

Citizen surveillance is embedded in various governance systems where citizens have access to information considered of interest and public domain. For this, the citizen or party interested in auditing or studying a particular issue seeks to ensure the data's authenticity and that it has not been deliberately altered in the process. Various technological solutions have been developed to verify data integrity, with entities like NIST having generated several current standards used within the Internet. These algorithms are widely used in secure transactions within digital systems, and require best practices and technological management policies to mitigate risks

As a result, security processes are gaining importance in a world where virtuality, remote operations, and data verification are becoming increasingly significant. Several centralized solutions have taken control of secure processes, where centralized systems are most commonly used in processes that guarantee privacy, authenticity, confidentiality, and integrity. However, most security failures are due to the fact that these centralized systems present single points of failure where critical operations can be compromised due to a security failure.[20]

Alternatives like blockchain, which decentralize some processes, have gained importance in recent years, mainly driven by concepts like digital asset ownership (NFTs) and cryptocurrencies. However, it is a specific case of DLT technologies, where information is open and nodes determine the validity of the added information through consensus algorithms. These decentralized mechanisms help to address some of the deficiencies of centralized systems and various implementations have been developed aimed at business process automation and certification [21]. Nevertheless, they also stand out for their high degree of energy inefficiency in most implementations, generating a significant cost to environmental sustainability given the energy resources required for validation and data addition within the network of nodes that make up the DLT [22].

Under these preliminaries, the question arises of how and whether citizens can trust in the integrity and authenticity of the data. This chapter presents various applications within the scope of a smart city, presenting an evaluation of the various algorithms used in DLTs, and a proposal for a secure electronic public document supported by biometrics, which allows for the authentication and integrity of public citizen transactions, reducing the risks of impersonation. This study also presents the proposal of a verification algorithm that allows for efficient implementation in computational and energy resources for a DLT system.

The second proposal, based on the results presented in the first proposal, revisits the analysis of the case study presented in this thesis, concerning how to efficiently integrate DLT technologies in the context of IoT, to ensure that measured data remains immutable and accessible to the citizen. This is achieved through a hybrid storage approach that leverages the benefits of both centralized and distributed systems. This allows the party interested in verifying the information to maintain control over data immutability by using interlaced hashes for data checking.

1.4.6 Other IoT Applications

Finally, through this thesis, the aforementioned chapters have addressed various scenarios focusing on well-being and air quality. Additionally, IoT can also be applied in various areas of human development. Once again, referring to the United Nations SDGs, Goal 4 aims to ensure inclusive and equitable quality education. In this goal, and in relation to the question of whether IoT can be employed to explore other aspects of human well-being, specifically regarding mental or productive focus, this chapter proposes an approach where IoT and non-invasive data collection can contribute in this aspect.

Particularly, the level of student attention is taken as a factor to analyze. This factor is gaining increasing attention due to the prevalence of distractions, such as electronic devices like mobile phones and computers being used for unrelated tasks during class, or simply emotional states. Coupled with the difficulty teachers face in analyzing feedback when dealing with large groups of students or when teaching remotely through learning platforms.

In this chapter, a system is presented that utilizes facial recognition and non-invasive physiological measurements to determine the level of student attention in

both remote and in-person classes. The results highlight the potential of combining these two approaches to enhance the effectiveness of teaching.

1.5 Thesis Contributions

Summarizing, this thesis presents the following contributions:

- An IoT system for the quantification of human well-being parameters in indoor and outdoor environments of a building
- An appropriate sampling rate that balances data accuracy and energy efficiency duty-cycle has been proposed, without sacrificing the quality of particulate matter concentration acquired by low-cost sensors.
- A novel hybrid model that integrates diverse deployment scenarios using a unified architectural framework and a standardized particulate matter monitoring station. This model harnesses a range of low-cost wireless technologies, thereby achieving high spatio-temporal resolutions.
- A framework to improve data quality, thereby increasing the coherence and reliability of IoT platforms that use low-cost light scattering sensors for air pollution monitoring
- A Blockchain network architecture proposal for a national e-ID system with iris and fingerprint recognition features. The design, implementation, and validation of a Blockchain network for the proposed e-ID system through a new consensus method called tournament consensus algorithm (TCA).
- A hybrid architecture for data storage that utilizes both centralized and decentralized storage resources, with data integrity safeguarded through the use of interlaced hashes.
- A novel IoT system is proposed, which utilizes facial expression recognition and physiological data analysis to assist teachers in evaluating teaching effectiveness by determining students' level of attention during both in-person and remote lectures.

Chapter 2

IoT Environments' Domain

In line with the Sustainable Development Goals (SDGs) set by the United Nations on September 2015 [23], the concept of smart cities has emerged as a strategic approach to foster sustainable urban development while addressing pressing societal challenges. The integration of Internet of Things (IoT) technologies, information and communication technologies (ICT), and data processing presents new opportunities to comprehensively monitor and manage urban environments.

This chapter explores the application of IoT technologies in two case monitoring studios: indoor monitoring in offices and outdoor monitoring air quality in the context of designing Healthy Cities. By leveraging IoT solutions, these applications contribute to the achievement of the SDGs by promoting sustainable and inclusive urban environments that prioritize the well-being and productivity of residents while addressing environmental concerns.

Indoor Environmental Quality Monitoring

The assessment of Indoor Environmental Quality (IEQ) in office spaces is crucial for creating healthier, more comfortable, and productive work environments. This section focuses on the development and characterization of a low-cost multi-sensor device known as PROMET&O (PROactive Monitoring for indoor EnvironmenTal quality & cOmfort). PROMET&O integrates a range of low-cost sensors to measure parameters such as air temperature, relative humidity, illuminance, sound pressure level, carbon monoxide, carbon dioxide, particulate matter, formaldehyde, and

nitrogen dioxide. This section delves into the calibration procedures employed to ensure reliable measurements, the integration of sensors within the device, and the usability of the collected data for different user groups, including building infrastructure managers and occupants.

Outdoor Environmental Quality Monitoring

The COVID-19 pandemic has brought to the forefront the importance of healthier urban environments and the need to monitor and manage environmental factors that impact public health. This section explores the role of IoT technologies in monitoring air quality on a large scale to support the development of Healthy Cities. The chapter presents a real-world case study where IoT-based Particulate Matter (PM) sensors, specifically targeting PM2.5 and PM10 particles, were deployed and compared to data from the official meteorological municipal station. The results highlight the benefits and challenges of implementing a comprehensive urban monitoring infrastructure, shedding light on the potential of IoT technologies for creating healthier and more sustainable urban environments.

By examining the application of IoT technologies in IEQ assessment and air quality monitoring, this chapter sets the stage for understanding the potential of IoT-based solutions in the context of smart cities. The insights gained from these applications offer valuable lessons and considerations for designing and implementing IoT-driven initiatives to create healthier, more sustainable, and livable urban environments.

Some of the work described in this chapter has been previously published in [24],[25] and some are currently being reviewed at an international conference on the topic.

2.1 Indoor Environmental Quality Monitoring

In recent times, there has been a surge of interest in the field of Indoor Environmental Quality (IEQ) due to the recognition that individuals spend a significant portion of their time indoors, particularly in office environments. The assessment of IEQ in offices is crucial as it encompasses several key factors such as thermal conditions,

lighting, acoustics, and indoor air quality (IAQ), all of which have a substantial impact on the comfort, well-being, health, and productivity of occupants [26, 27].

Consequently, there is a growing demand for methodological approaches that can comprehensively evaluate IEQ conditions by considering these domains in a holistic manner rather than treating them as isolated elements. The assessment of IEQ involves on-site monitoring of parameters and indexes related to these domains, and recent advancements have facilitated intensive and long-term monitoring campaigns through the use of cost-effective sensors, wireless sensor networks, and cloud-based software platforms [28, 29].

As example Tiele et al.[28] developed a monitoring device for assessing IEQ in indoor working environments. The device is capable of monitoring various parameters including temperature, humidity, illuminance, sound levels, particulate matter, total volatile organic compounds, carbon dioxide, and carbon monoxide. The sampling period for data collection is set at 10 minutes. This device is designed to be low-cost and operates on battery power, with an estimated battery life of 68 hours. Its enclosure dimensions measure $165mm \times 105mm \times 55mm$. To ensure accurate measurements, the temperature, humidity, and carbon dioxide sensors of the device were calibrated using the CO210 Extech commercial system. It was observed that the temperature and carbon dioxide readings deviated from the uncertainty values provided by the manufacturers. As a result, adjustments were made to the temperature reading by 1.9 °C and to the carbon dioxide reading by 70 ppm. Furthermore, to establish the baseline characteristics of the IEQ device, it was placed inside a sealed plastic enclosure along with the CO210 Extech and exposed to zero air. This step aimed to gain insights into the device's performance under controlled conditions. Additionally, the researchers devised a rating system for the collected measurements, categorizing each parameter as *good*, *average*, *poor*, or *bad* based on the extent to which they surpass predefined thresholds. Moreover, they calculated an IEQ index by taking a weighted average of the individual parameters, and the resulting percentage value is displayed on a low-power OLED screen incorporated into the device's casing.

Parkinson et al.[29] also developed a monitoring device for assessing IEQ specifically designed to be placed on office desks. Known as SAMBA, this low-cost device is equipped to measure various parameters including air temperature, relative humidity, globe temperature, air velocity, sound pressure level, illuminance, total volatile

organic compounds, formaldehyde, carbon monoxide, and carbon dioxide. SAMBA consists of two separate units connected by an ethernet cable. To ensure accurate temperature measurements, the temperature-sensitive transducers are housed in the satellite unit, isolated from heat-emitting components. The thermal sensors undergo calibration in a small-scale wind tunnel. For the calibration of indoor air quality sensors, a sealed chamber is utilized, where reference gases are introduced through an intake port, and the sensors and reference devices monitor the concentration. The calibration process for illuminance sensors involves placing a dome on top of SAMBA and using an RGB LED module (WS2812, Worldsemi) as a point-source controlled via PWM. In order to calibrate the microphone, it is positioned near a monitor generating a noise signal within the frequency range of 100 Hz to 16000 kHz, alongside a reference sound pressure level (SPL) meter (Type 1; NL-52, Rion). They also created a user-friendly software platform for reporting the monitored data. The reported IEQ score is calculated based on hourly averages of the measured parameters and their compliance with international standards.

To date, the monitoring of IEQ has primarily focused on specific research questions such as its correlation with energy consumption or occupants' comfort and behavior. However, there is a need for a comprehensive strategy that encompasses all these aspects. This is because IEQ has a significant impact on Indoor Environmental Comfort (IEC), which relates to the overall sense of well-being and satisfaction experienced by individuals. The advancement of information and communication technology enables the continuous collection of occupants' subjective feedback through portable computers or smartphones. Nevertheless, there is no universally recognized standardized solution regarding the number, type, and frequency of questions required for this purpose [30]. Additionally, various contextual, personal, and behavioral variables influence occupants' perception of their environment, introducing potential biases in their responses [31]. These variables may include factors such as building orientation, workstation location, age, country of birth, and educational qualifications. It is worth noting that occupants' dissatisfaction with the environment can persist even when the physical conditions meet standard requirements. This dissatisfaction can subsequently affect their behavior and lead to consequences on building energy consumption [32].

The PROMET&O device (PROactive Monitoring for indoor EnvironmenTal quality & cOmfort) is designed as a comprehensive system for continuous in-field monitoring of IEQ parameters using multiple sensors. Its intended application

includes integration with Building Automation and Control Systems (BACS). Additionally, the device incorporates a tailored questionnaire to capture occupants' subjective feedback regarding their comfort perception. This feature enables the correlation of objective and subjective data. Ensuring the reliability of the monitored data is crucial, which involves verifying the uncertainty specifications provided by manufacturers and conducting calibration procedures as necessary.

2.1.1 Indoor Environmental Monitoring Quality Requirements

One aspect of this research is the design and metrological characterization of the PROMET&O multi-sensor device. Currently, there is no universally recognized set of parameters and indexes for assessing Indoor Environmental Quality (IEQ) [33], so the standards for individual domains serve as the primary reference. Table 2.1 provides an overview of the parameters incorporated in the multi-sensor device and their respective thresholds for office environments.

2.1.2 PROMET&O System

The primary objective of the PROMET&O system is to promote a proactive approach among building occupants while simultaneously monitoring the IEQ conditions and IEC. The system consists of several key components, as illustrated in Figure 2.1. Each multi-sensor device is responsible for measuring various parameters related to thermal, lighting, acoustic, and air quality conditions, and transmits pre-processed data to a cloud-based platform. All objective data are securely stored in the cloud and can be accessed by users through an intuitive dashboard. Occupants also have the opportunity to provide subjective feedback on their perceived IEQ through a questionnaire. By integrating objective IEQ data with subjective IEC data, collected from user feedback, the system aims to foster a proactive behavior among occupants. A comprehensive description of each component is presented in the subsequent subsections.

Multi-sensor Device

The design of the PROMET&O multi-sensor has been guided by the measurement requirements outlined in the subsection 2.1.1 and Table 2.1. This low-cost device

Table 2.1 Parameters under monitoring and threshold values according to international standards and the WELL Building Standard

Parameter	**Threshold for Offices**	**Reference**
Air temperature (T)	Winter: (20-24) °C Summer: (23-26) °C	ISO 7730:2005
Relative Humidity (RH)	(25-60) %	EN 16798-1:2019
Vertical Illuminance (E_v)	Writing, typing, reading $\geq 500lx$	EN 12464-1:2021
Sound Pressure Level (SPL)	$\geq 45dB(A)$	NF S 31-080
Carbon monoxide CO	15 min. mean $\geq 100mg/m^3$ 1 h mean $\geq 35mg/m^3$ 8h mean $\geq 10mg/m^3$ 24 h mean $\geq 7mg/m^3$	EN 16798-1:2019
Carbon dioxide (δCO_2)	$\geq 800ppm$	EN 16798-1:2019
Nitrogen dioxide (NO_2)	1 h mean $\geq 200\mu g/m^3$ Annual mean $\geq 20\mu g/m^3$	EN 16798-1:2019
Particulate matter ($PM_{2.5}$)	24 h mean $\geq 25\mu g/m^3$ Annual mean $\geq 10\mu g/m^3$	EN 16798-1:2019
Particulate matter (PM_{10})	24 h mean $\geq 50\mu g/m^3$ Annual mean $\geq 20\mu g/m^3$	EN 16798-1:2019
Formaldehyde (CH_2O)	30 min. mean $\geq 100\mu g/m^3$	EN 16798-1:2019
Total volatile organic compounds	$\geq 500\mu g/m^3$	WELL

Table 2.2 Sensors References

Parameter	**Part Number**	**Manufacturer**	**Interface**
T and RH	SHT41	Sensirion	I2C
E_v	VEML7700	Vishay	I2C
CO	3SP_CO_1000	Spec sensors	Analog
CO_2	SCD30	Sensirion	I2C
NO_2	3SP_NO2_5FP	Spec sensors	Analog
$PM_{2.5}$ *and* PM_{10}	SEN54	Sensirion	I2C
CH_2O	SFA30	Sensirion	UART
SPL	IMP34DT05	ST	I2S

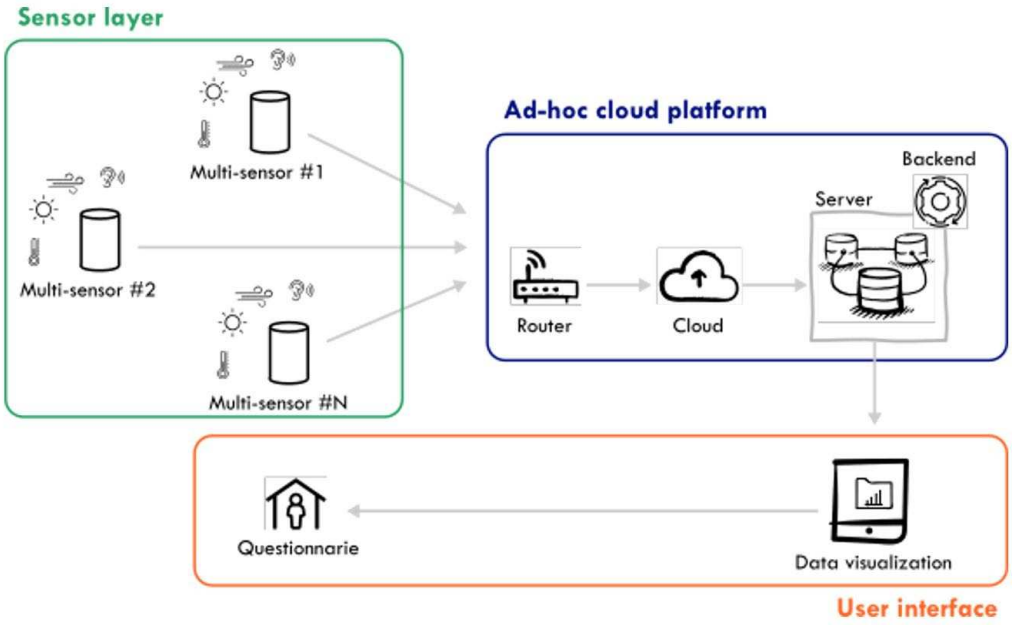


Fig. 2.1 PROMET&O System Architecture

utilizes off-the-shelf sensors and is structured around three main components (see Figure 2.2): the sensor array, the peripheral interface and power circuitry, and a micro-controller responsible for acquiring sensor outputs and implementing calibration functions for each measuring chain. Additionally, a WiFi module is incorporated to facilitate the transmission of acquired data to a cloud platform for visualization, post-processing, and storage. The communication between the microcontroller, the different sensors, and the WiFi communication module is established through the PCB, as represented by the peripheral interface circuitry block. Additionally, the device's various sub-blocks are powered by low voltage DC-DC converters connected to an external power adapter to minimize self-heating concerns. All these components are housed within a single-case device, with embedded sensors enabling the measurement of the quantities specified in Table 2.1. To mitigate the impact of self-heating from the micro-controller and WiFi module (with a power dissipation of approximately 300 mW), the sensor array is thermally isolated from the rest of the system. The internal volume of the device is divided by a panel, providing mechanical support and separation between the sensor compartment and the controller. Furthermore, the multi-sensor features a PLA case that encompasses all the electronics.

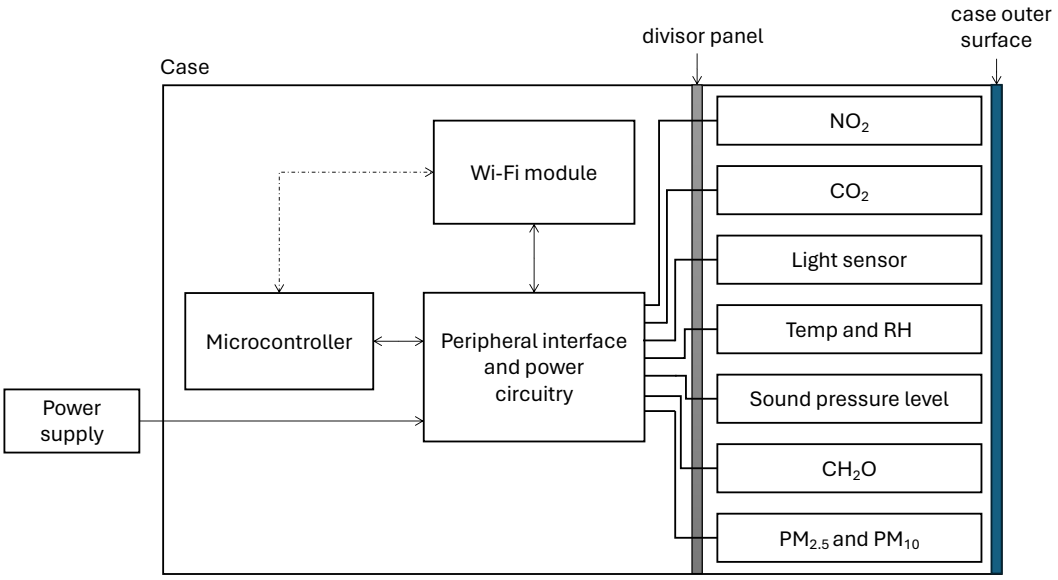


Fig. 2.2 Block scheme of the multi-sensor device

Figure 2.3 reports a photograph of the multi-device prototype looking at the sensors compartment and the micro-controller position. The position of each sensor inside the case has been carefully chosen to avoid cross-interference. The air-quality sensors are placed on an ad-hoc vertical mount to get the sensitive elements as close as possible to the case drilling. In such a way, their measurements are expected to be representative of the indoor environment to be monitored. On the contrary, the illuminance sensor and the microphone are placed on the top of the device to negligibly affect their spatial responses.

The external enclosure of the multi-sensor device is made of PA12 material and has a cylindrical shape, measuring 18 cm in height and 12 cm in diameter. Figure 2.4 shows a photograph of the case. The top cover of the case has a slightly smaller perimeter, creating a gap in the ring to facilitate heat dissipation from the operating sensors. The illuminance sensor and the MEMS microphone are positioned on the top cover as well. To ensure proper ventilation and prevent component overheating, the case has a series of holes on both the front and back sides along the perimeter. Additionally, there are ten openings on each side above the holes, housing ten LEDs that indicate the calculated percentage of IEQ based on specific algorithms considering the monitored parameters. At the base of the case, there are two larger openings allowing the $PM_{2.5}$ and PM_{10} sensors to draw in and expel air using a built-in fan. For power supply connections, two jacks are provided—one on the side

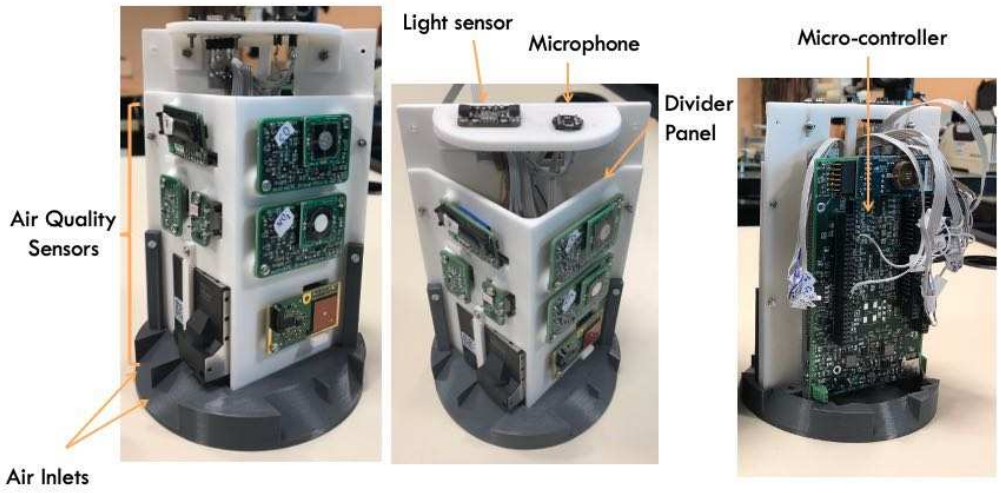


Fig. 2.3 Placement of the PROMET&O sub-blocks

for desk installation and another at the bottom for pole-mounted or wall-mounted installation.

The microcontroller is required to support two I2C interfaces (one for sensors and one for the WiFi module), two UART interfaces (one for the formaldehyde sensor and one for debugging), two ADCs for analog sensors, SDRAM memory, and several available pins for LED connections.

The inclusion of memory depends on the quantity of measurements taken by the sensors that the system needs to retain for internal processing before transmitting results to the server. The measurement that has the greatest impact on memory usage is audio, with a sampling rate of 22 kHz (22,000 samples per second) at 16 bits, considering a 5-second acquisition time, for example. This calculation results in 214 kB of occupied memory. These requirements have led to the selection of STMicroelectronics' F7 and H7 families, both with a minimum of 176 pins. The chosen controller board is the NUCLEO-H7A3ZI-Q, which includes a microcontroller from H7 family.

The selection of sensors was based on the referenced documents in Table 2.1 and 2.2. The selection criteria took into account factors such as range, measurement uncertainty, cost, power consumption, physical dimensions, and response time. Temperature and relative humidity measurements are provided by a digital ultra-low power sensor. The illuminance sensor used is a photodiode with a spectral response that closely matches the sensitivity of the human eye in daylight conditions. The

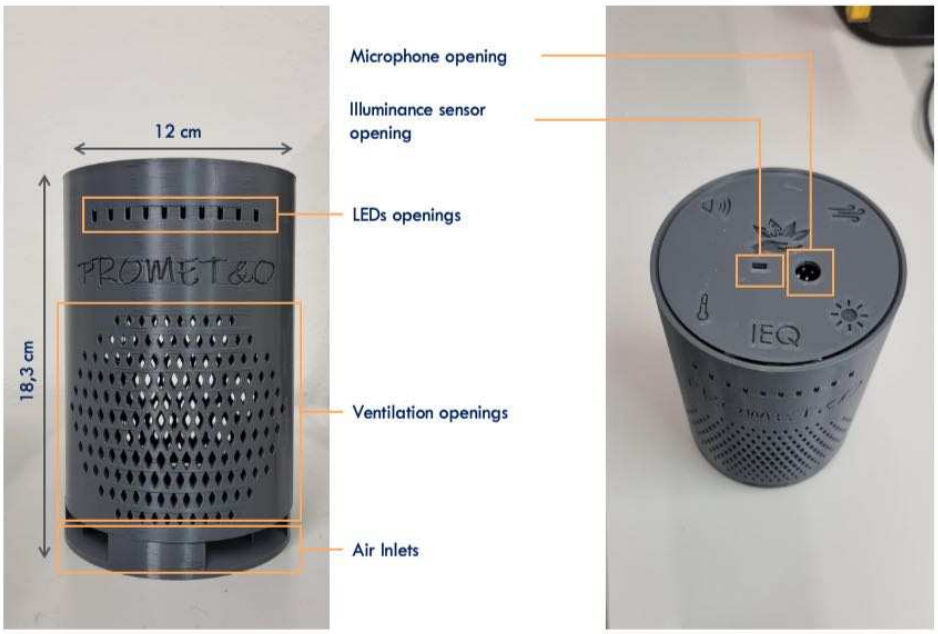


Fig. 2.4 Photograph of the external case

microphone employed is an omni-directional MEMS microphone that utilizes a capacitive sensing element. Nitrogen dioxide and formaldehyde sensors rely on electrochemical cells for detection. The carbon dioxide sensor utilizes Non-Dispersive InfraRed (NDIR) technology for measurement, and the particulate matter sensor employs an optical particle counter (OPC) based in light-scattering for detection.

The assurance of measurement traceability for the designed multi-sensor device necessitates defining appropriate calibration procedures for each quantity of interest within the measurement chain. The metrological characteristics of the selected sensors determine two distinct conditions for each quantity. If the sensor manufacturer's stated uncertainty meets the requirements, a verification procedure is conducted for the entire measurement chain, comparing it against a reference standard to assess the overall error. On the other hand, if the sensor specifications do not meet the uncertainty requirements, a metrological characterization is necessary to modify the calibration function of the measurement chain using a reference standard. These two conditions are visually represented in Table 2.3, with different background colors in the *Uncertainty* column: white indicating sensors capable of meeting uncertainty requirements (first condition), and gray indicating sensors requiring appropriate characterization (second condition). The required uncertainty values are sourced from

Table 2.3 Main characteristics of the selected commercial sensors

Parameter	**Sensor Measurement Range**	**Uncertainty**	**Range**	**Required Uncertainty**
T	(-40 – 125)°C	± 0.2 °C	(0 – 60)°C	±0.5°C (BS EN ISO 7726:2001)
RH	(0 – 100)%)	± 1.8%)	(30 – 70)%)	±5% (ANSI/ASHRAE 55:2017)
E_v	(0 – 120) (klx)	15% (mv†)	-	±5% (WELL)
CO	(0 – 1000) ppm	± 2.75 nA/ppm (sensitivity)	-	1 ppm at values $\sim$ (0 – 10) ppm (WELL)
CO_2	(0 – 40000) ppm	± (30+3% mv†)	(400 –10000) ppm	10% at 750 (WELL)
NO_2	(0 – 5) ppm	±30 % mv†	(0 – 5) ppm	20% (WELL)
$PM_{2.5}$	(0 – 1000) $\mu g/m^3$	± (5 $\mu g/m^3$ + 5% mv†)	(0-100) $\mu g/m^3$	$\geq$15% (WELL)
PM_{10}	(0 – 1000) $\mu g/m^3$	±25 $\mu g/m^3$	(0 – 100) $\mu g/m^3$	-
CH_2O	(0 – 1) ppm	±20 % mv†	(0-200) ppb	20 ppb (0-100 ppb) (WELL)
SPL	122.5 dB (SPL) AOP	Not declared	-	±0.5 dB (1 kHz) (WELL)

† mv: measured value.

International Standards or building certification schemes, and the specific procedures for different quantities are described in subsection 2.1.3.

Ad-hoc cloud platform

The multi-sensor devices send the IEQ data at intervals as short as 1 second, necessitating the adoption of a lightweight communication protocol. To ensure the reliable and efficient transmission of the measured data on Indoor Environmental Quality (IEQ), the PROMET&O system utilizes the Message Queue Telemetry Transport (MQTT) protocol. This protocol follows a publish-subscribe paradigm and offers payload optimization capabilities. Within the system's back end, a broker and a client are employed to collect data from the multi-sensor devices and store it in a database similar to MySQL. Subsequently, the system's front end retrieves the data from the database and presents it to users through a user-friendly dashboard.

User Interface

The primary objective of the developed dashboard (see Figure 2.5) is to provide end-users with real-time and historical information about the physical conditions of their environment. It aims to enhance users' knowledge of IEQ through the HINTS and MORE sections, thereby increasing their awareness and understanding of their role in improving their IEC. Additionally, the implemented questionnaire allows users to express their perception of IEC for each specific domain. The questionnaire also collects personal and behavioral variables to enable the analysis and comparison of IEC with IEQ.

To validate the effectiveness of the dashboard and questionnaire, two campaigns were conducted. A panel of subjects was presented with closed questions using Microsoft Form and interviews to evaluate the ease of navigation, comprehension of content, and quality of the graphical layout. This assessment aimed to enhance user engagement and ensure the overall quality of the dashboard and questionnaire.

Fig. 2.5 Dashboard screenshot. It provides a quick overview of the current IEQ for the user.

2.1.3 Calibration Procedures

Air Temperature and Relative Humidity

The validation procedure for the temperature and humidity sensor(*Sensirion SHT41*) involves placing the device inside a climatic chamber and comparing its readings to those of a reference thermo-hygrometer. The reference instrument ensures an uncertainty of 0.1 °C for temperature and 1.5% for relative humidity within the desired ranges. Additionally, a Pt100 is used as an additional reference, following the recommendations of E220-19 [34].

For temperature verification, tests are conducted at three different set points (10 °C, 20 °C, and 30 °C) with a sampling interval of 60 seconds. Data collection begins when the climatic chamber reaches the set temperature of 10 °C. The conditions are then manually adjusted to the next set-point without stopping the recording, allowing for measurements in both steady-state and transient conditions.

Regarding the verification of relative humidity measurements, the temperature is set at 23 °C, and four different relative humidity values (22%, 39%, 75%, and 94%) are tested. The same 60-second sampling interval is used for data collection from both the sensor under verification and the reference device. Data acquisition starts

once the required thermo-hygrometric conditions are reached and continues for a time interval longer than 1 hour for each test point.

Sound Pressure Level

The verification of the MEMS microphone is conducted both as an individual sensor and as an integrated component within the multi-sensor case to consider the impact of acoustic diffraction effects caused by the mounting configuration. A secondary free-field calibration is performed using the comparison method. The free-field calibration based on the comparison method follows the guidelines of the IEC 61094-8 Standard and is carried out in either an anechoic or hemi-anechoic chamber (refer to Figure 2.6). In this method, the free-field sensitivity of the microphone under test is determined by comparing it to a reference microphone with known free-field sensitivity. (The reference microphone's sensitivity is obtained through primary free-field reciprocity calibration or derived from primary pressure reciprocity calibration with appropriate free-field corrections). Both microphones are exposed to essentially the same free-field sound pressure in a sequential manner during the calibration process.

The monitoring of the sound source's stability and potential variations in the acoustic field during measurements is ensured by a fixed-position monitor microphone placed inside the chamber. Additionally, the acoustic pressure is measured in a specific region of the test environment where sound generated by the loudspeaker propagates in a nearly planar progressive wave manner. Both the separate and case-integrated MEMS microphones are positioned at the same measurement point, following the specified angle of incidence, to establish their reference points or acoustic centers. The free-field sensitivities of both microphones are determined through comparison or substitution calibration within the frequency range of 500 Hz to 12.5 kHz. This evaluation is conducted in a small anechoic chamber with a volume of approximately $3.5m^3$.

Illuminance

The calibration process for the illuminance measurement chain involves comparing it to a reference standard, specifically the *PRC Radiolux 111* luxmeter, which features a photometric head certified as class B according to DIN 5032-7. The calibration

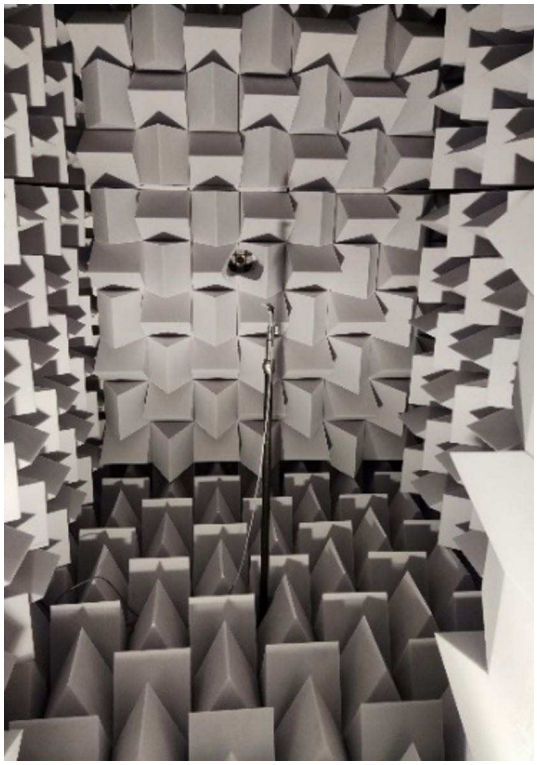


Fig. 2.6 The anechoic chamber used for the MEMS microphone calibration

method consists of measuring the illuminance within a test box using a stable light source that can be dimmed to various light output levels, as shown in Figure 2.7. In the first step, the reference standard measures the illuminance as a baseline, and in the second step, it is replaced with the sensor being calibrated to obtain its readings under each test condition. The calibration method encompasses the following test conditions:

- Three LEDs with different spectra and Correlated Color Temperatures of 2700 K, 4000 K, and 5700 K (representing warm white, neutral white, and cool white, respectively) are utilized as light sources in the test box to evaluate the spectral response of the sensor being calibrated.
- Each LED is dimmed to achieve illuminance values ranging from 2.5 lx to approximately 3500 lx on the measuring plane inside the test box.
- The sensor under calibration is positioned horizontally and tilted at different angles (30° and 60°) to assess its cosine response to varying angles of light incidence.

- The sensor under calibration is tested both with and without the PROMET&O device case, considering different window openings and cover thicknesses applied to the case.

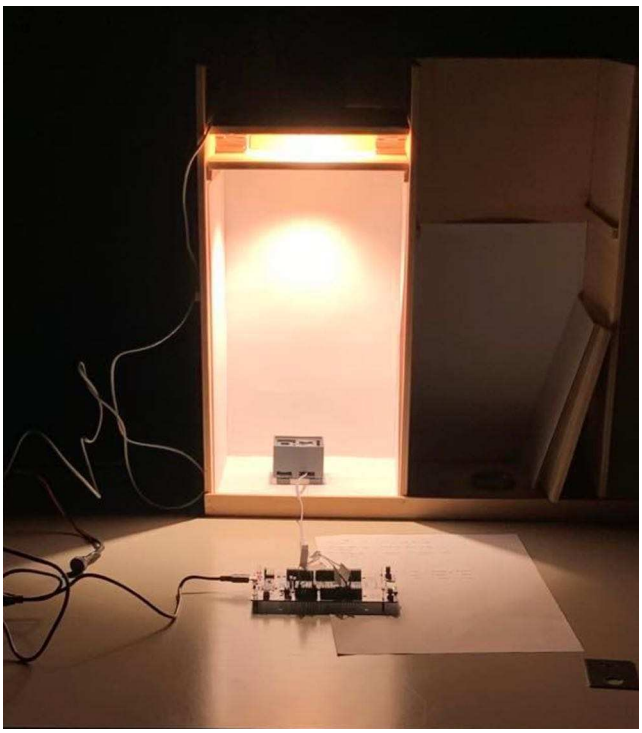


Fig. 2.7 Data acquisition phase of the naked sensor with LED 2700 K without optical filter

Throughout all the tests, the distance between the light source and the sensitive area remains constant to ensure a consistent and uniform distribution of light on the measuring plane. The accurate positioning of the sensitive area beneath the light source is verified using a laser beam pointer. The test box houses the LED source, and the measuring plane is located in a completely darkened room with stable thermal conditions. The tests are conducted in a random manner, and the reproducibility of the results is assessed to evaluate the consistency of the measurements.

Particulate Matter

Low-cost light-scattering PM sensors are compact versions of traditional optical particle counters, designed for integration into IoT solutions [35, 36]. Although calibration and verification methods for light-scattering airborne particle counters

(LSAPC) are defined in ISO 21501-4:2018 [37], many low-cost sensors do not adhere to this standard. These sensors operate by illuminating an air sample with a laser beam and measuring the intensity of scattered light using a photodiode positioned at a specific angle relative to the laser source. Utilizing Mie theory, the particle diameters can be determined, and the number of particles within various size intervals can be counted. The particle volume is then calculated by assuming an ideal spherical shape [38]. From the total volume, the total mass of particulate matter (PM) is estimated, considering a density coefficient that may vary across different size intervals. The PM concentration can be obtained by dividing the total mass by the volume of the analyzed air sample. It should be noted that this measurement procedure assumes spherical particles, which is not representative of environmental pollution. Additionally, the density coefficient used in factory calibration may not accurately reflect the real working environment. Therefore, it is beneficial to recalibrate the sensor using PM particles similar to those encountered in the deployment scenario. The literature presents various models for calibration [16, 39, 13], with this work utilizing a simple linear regression model to correct for significant errors resulting from factory calibration. Since sensor calibration is strongly influenced by the composition of particulate matter, two types of particulates are considered: cigarette smoke and outdoor pollution in the installation area. The calibration for cigarette smoke is conducted in a sealed climatic chamber against a high-precision reference sensor, while for outdoor monitoring, the sensors are positioned near official monitoring stations. Additional adjustments can be introduced to the model by incorporating temperature and humidity measurements. Initially the sensors are calibrated by the manufacturer under the technique described above, so in the indoor case this process is not performed for this monitoring station.

2.1.4 Experimental Results

Initial instances of experimental outcomes are presented for two of the measurement systems under investigation, specifically targeting the measurement of CO_2 concentration and illuminance (E_v). In accordance with the classification outlined in subsection 2.1.2, a validation procedure is conducted for the CO_2 concentration measurement, whereas a comprehensive characterization of the E_v measurement system is warranted.

The findings pertaining to the validation of the CO_2 concentration measurement system are summarized in Figure 2.8 for the reference set-point (approximately 500 ppm). The upper chart displays the reference concentration as provided by the Photoacoustic Gas Monitor (PGM, depicted by the green line), alongside the concentration measured by the *Sensirion SCD30* device (indicated by the red line) over a duration of approximately 70 minutes. The lower chart presents the measurement error (blue line), representing the disparity between the readings of the sensor under verification and the PGM. Additionally, the chart includes the upper limit (UL SCD30) and lower limit (LL SCD30) of the acceptable error range for the SCD30 sensor (depicted by the red lines), as well as the confidence interval of the measurement error (represented by the green lines) at a 95% confidence level. The expanded uncertainty U(Error) has been calculated, taking into account the contributions stemming from the PGM uncertainty and the resolution of the sensor under verification. It is worth noting that the measurement error falls within the specified tolerance range, indicating conformity of the CO_2 measurement system under verification to its nominal specifications. A comprehensive statement of conformity will be provided upon completion of the verification procedure across the entire range of interest.

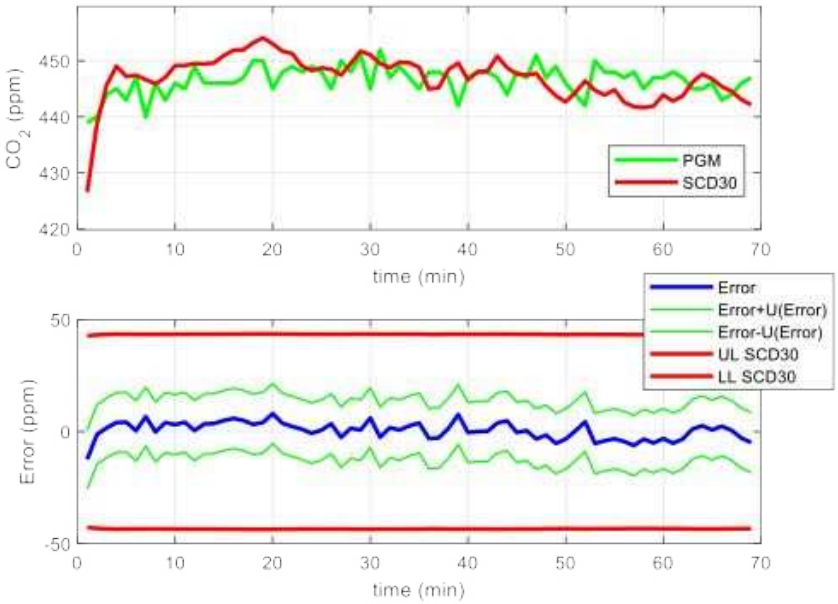


Fig. 2.8 Verification results obtained for the CO_2 concentration measurement chain in the baseline set-point.

Regarding the illuminance measurement system (E_v), an initial characterization is necessary due to the manufacturer's specified maximum permissible error (± 15% of the measured value). Consequently, the calibration function of the illuminance measurement chain was assessed using the results obtained from the LED with a correlated color temperature of 2700 K (warm white). This LED possesses a spectral response that closely aligns with the peak sensitivity of the human eye in photopic conditions ($V(\lambda)$). The findings are summarized in Figure 2.9, where the red circles in the upper chart represent the experimental values (reference vs. measured), and the blue line represents the linear calibration function derived by minimizing the root square sum of the differences between the function and the experimental values. The linear calibration function is defined by an intercept of -35 lx and a slope of 1.22 lx/lx.

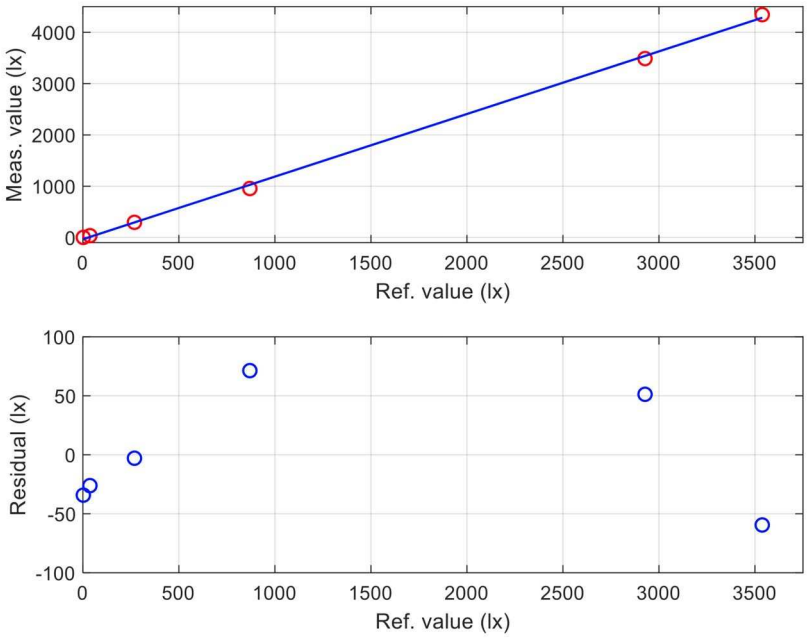


Fig. 2.9 Characterization results for the illuminance measurement chain obtained using the LED with the correlated color temperature of 2700 K

The residual fitting errors (shown in Figure 2.9 bottom chart) exhibit a minimal average value, and the root mean square error measures approximately 50 lx. This contribution to the uncertainty has been combined with the uncertainty stemming from the reference luxmeter, resulting in the anticipated uncertainty of the characterized illuminance measurement chain. The uncertainty value, expressed with a 95% confidence level, can be described as follows:

$$U_{adj}(E_v) = (60 + 4\%measured\ value)lx \tag{2.1}$$

Subsequently, the illuminance measurement chain underwent verification by comparing it to the same reference device using the remaining two LEDs with correlated color temperatures of 4000 K and 5700 K (which have maximum sensitivity in the violet region of visible light), as well as the 2700 K LED after a month from the initial characterization. The results obtained from these comparisons are presented in Figure 2.10, where the red symbols represent the errors obtained using the unadjusted indications of the illuminance measurement chain, while the blue symbols represent the errors resulting from implementing the identified calibration function. Additionally, the continuous red lines in the figure represent the expanded measurement uncertainty of the characterized chain. The key finding of this analysis is the efficacy of the proposed characterization procedure, which enables the measurement of illuminance (E_v) with an acceptable level of uncertainty, meeting the requirement outlined in Table 2.3. Conversely, when the illuminance sensor is not adjusted, the measurement error exceeds the specified uncertainty (±15% of the measured value), as evident from the unadjusted errors observed at illuminance values exceeding 3500 lx.

2.2 Outdoor Environmental Quality Monitoring

2.2.1 Introduction and Background

The urban environment significantly impacts human health, with Atmospheric Particulate Matter (PM), nitrogen dioxide, and tropospheric ozone being recognized as risk factors for cardiovascular, respiratory, and carcinogenic diseases [40]. Alarming data from the scientific community reveals that approximately 4.2 million premature deaths occur worldwide each year due to stroke, heart disease, lung cancer, and chronic respiratory illnesses resulting from excessive exposure to PM [40]. In Europe, where 15% of premature deaths can be attributed to environmental factors, life expectancy has been reduced by about 8 months due to the same causes [41].

Against this backdrop, the COVID-19 pandemic has further emphasized the crucial link between environmental factors (such as virus spread and exacerbation

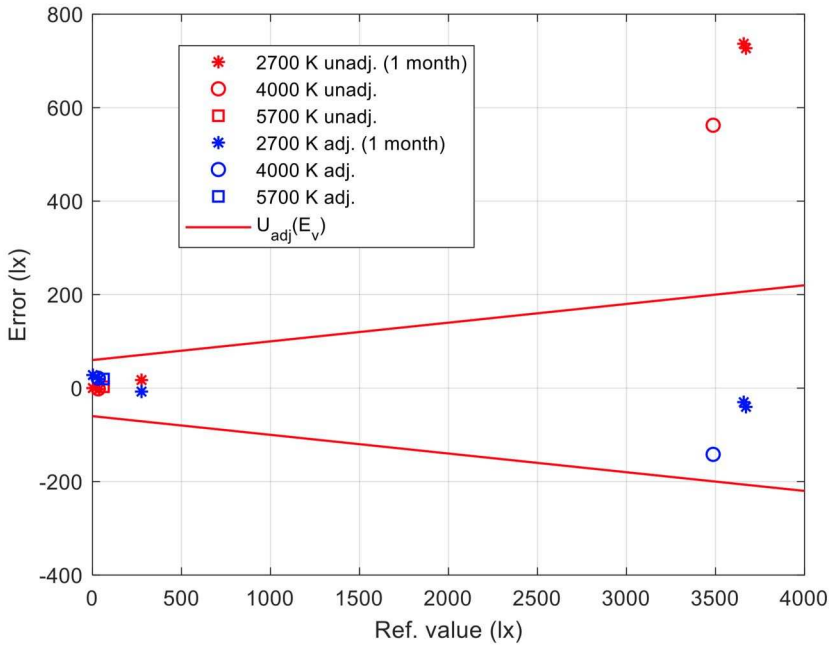


Fig. 2.10 Results of the illuminance measurement chain verification. The indications without any correction (unadjusted, represented by red symbols) and the indications after applying the identified calibration function (adjusted, represented by blue symbols).

of symptoms) and human health, underscoring the need to promote healthier urban environments [42]. The complexity of the factors affecting air quality, which operate at local, continental, and hemispheric scales, is influenced by human activities (e.g., heating, transportation, industries) and micro-climatic conditions that transform primary emissions into secondary pollutants, facilitating their dispersion [43].

In this situation, affordable ICT in conjunction with IoT solutions enable the creation of a digitally connected urban environment. On the one hand, these technologies can serve as essential tools for extensive monitoring, while on the other hand, they can effectively engage people by increasing awareness of health issues and promoting the transition to an ecological lifestyle [13, 44].

Numerous ongoing studies are exploring the use of a widespread network of air quality sensors integrated into various elements such as lampposts, vehicles, bicycles, building components, flowerpots, or wearable devices [45–48]. Specifically, the development and application of IoT technologies in cities allow for monitoring multiple variables that can impact air quality and human health. Additionally, adopting integrated technologies supports the growth of open urban platforms [49],

which monitor the inflow and outflow of energy and matter in real-time urban metabolism models [50].

Based on these foundations, this study presents initial findings from an air quality monitoring campaign conducted at the neighborhood level in Turin, located in northern Italy, which is considered one of the most critical regions in Europe due to its orography and population density. The study aims to achieve two primary objectives: firstly, to demonstrate the effectiveness of low-cost IoT technologies in monitoring air quality on a large scale, and secondly, to investigate the impact of the urban environment on pollutant distribution. The following subsections provide a description of the methodological approach employed and highlight the key results obtained through a test conducted as part of an ongoing national research project.

2.2.2 Methodology

This subsection outlines the experience of a comprehensive monitoring campaign focused on air quality ($PM_{2.5}$ and PM_{10}) and environmental parameters (such as air temperature, humidity, and pressure) conducted at a neighborhood level with a high level of temporal resolution. This micro-urban scale campaign aimed to explore the potential direct relationship between various built environment characteristics (e.g., morphology, presence of green spaces, street exposure, building height, orientation, etc.) and the distribution of pollutants. The findings have the potential to inform climate change planning policies and designs [51].

Prior to the monitoring campaign, the area was simulated using *ENVI-met*™, a Computational Fluid Dynamics software (CFD), to better understand the dispersion of pollutants in the micro-climate. Over the course of one year (from March 2021 to February 2022), three monitoring campaigns were conducted, each lasting approximately two weeks. Four monitoring stations were strategically placed in the area for data collection. The data collected from the low-cost sensor stations developed by Giusto et al. [13] were contrasted with data from the official urban measurement network operated by ARPA (Regional Environmental Protection Agency), the municipal authority responsible for urban air quality control. For comparison, data from the nearest official station, Torino Grassi – Reiss Romoli, located approximately 4 *km* away from the analyzed area, was utilized. Furthermore, the data analysis

considered different seasons and exposures to primary emission sources to examine the impact of seasonal anthropogenic phenomena.

2.2.3 Case Study

The Neighbourhood

The case study takes place in the northeastern suburbs of Turin, Italy, a city characterized by *Cfa* climate conditions according to the Köppen-Geiger classification. This climate type signifies a humid subtropical climate with the coldest month averaging above 0 °C, at least one month with an average temperature above 22 °C, and at least four months with an average temperature above 10 °C. The specific neighborhood under investigation is Milan Barrier (45° 05'N 7°42'E), situated near the Po River and several major green infrastructures.

The plot covers an area of approximately $22,000\ m^2$ and is predominantly occupied by public housing facilities (refer to Figure 2.11). To be more specific, the area comprises two building block courtyards, each consisting of three to four storeys in height. Those were constructed in the 1920s and 1940s and have undergone technological and functional renovations in 2012, including facade upgrades, installation of external lifts and new balconies, and the renewal of water systems. The residential complex is owned by ATC (Housing Territorial Agency in Piedmont Region), and it is currently home to around 200 people, with approximately 30% of the population aged 60 or above.

The selection of this particular plot was motivated by the diverse presence of natural and artificial elements, allowing for an assessment of how pollutants propagate within the urban fabric. Specifically, the choice of the closed-courtyard morphological type, commonly found in Turin's historical fabric, was made to examine potential variations in pollutant distribution between the facades facing the street (thus more exposed to vehicle emissions) and those facing the internal green area. The entire area is adjacent to an old railway path in the north, which is set to become part of a new metro line infrastructure, offering significant environmental potential. All the monitoring stations were installed on balconies at a similar height above the ground (approximately 6 meters) to ensure consistent pollution patterns and minimize any potential variations due to height differences.

Fig. 2.11 Aerial view of the case study neighborhood

CFD Simulation

Before installing the sensors, an outdoor Computational Fluid Dynamics (CFD) analysis was conducted on the area (see Figure 2.12). This analysis aimed to understand how ventilation, a critical factor in pollutant propagation, could influence the results. *ENVI-met*, a widely used CFD model in academic and professional fields, was employed to simulate the complex interactions between the built and natural environment by a prognostic non-hydrostatic model [52]. The modeling process involved defining various features of the built environment, such as aterial albedo, type of greenery, and soil moisture.

Additionally, meteorological boundary conditions, including temperature, humidity, wind, and solar radiation, were set to accurately represent the weather. The simulation provided insights into various outdoor parameters (e.g., Potential Air Temperature (PAT), Surface Temperature (SF), Wind Speed (WS), etc.) and thermo-hygrometric comfort indices (e.g., Physiological Equivalent Temperature (PET)).

The analysis of the case study area revealed that the linear region to the northeast of the plot, previously a railway yard, significantly enhanced ventilation. For the simulation, the initial wind speed and direction from ARPA for the hottest summer day in 2020 were used ($1.70\ km/h, NW$). Although simulation results represent specific microclimatic conditions, the wind and their direction are typical for Turin.

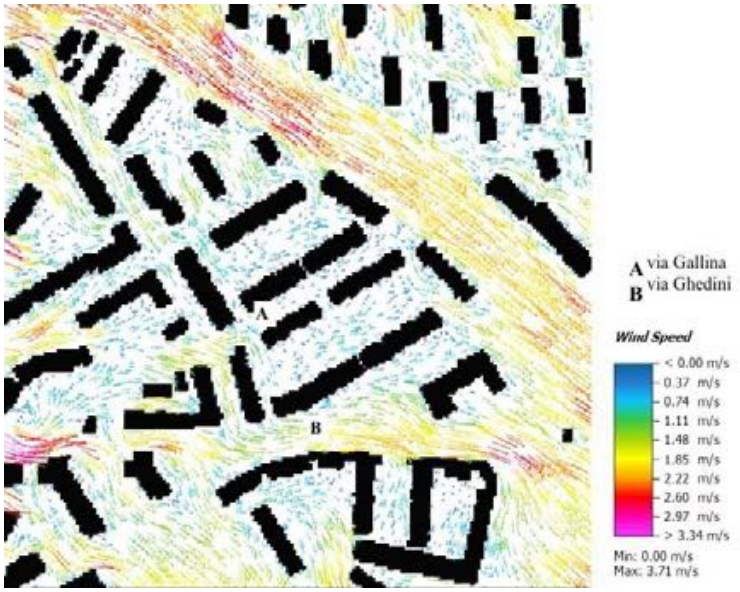


Fig. 2.12 ENVI-met's CFD analysis of the plot revealed the wind speed and direction at 2:00 PM on the hottest summer day

To examine the impact of wind on PM propagation, two sensors were positioned to face the street, with one sensor facing a more open road (via Ghedini) and the other facing a relatively closed road (via Gallina). The remaining two sensors were placed facing the courtyard, where lower wind speeds were anticipated based on the simulation outputs.

Monitoring Board Station

The monitoring results specifically include data collected from PM_{10} and $PM_{2.5}$ sensors, as well as temperature, humidity, and pressure measurements. The hardware components used for the monitoring board (Figure 2.13) are as follows:

Monitoring board: The Raspberry Pi Zero Wireless was selected as the single-board computer for the system. To optimize resources and minimize their wastage, the project employed the Arch Linux operating system for ARM, which focuses on running only essential OS components. Python scripts were utilized as System Units to query all the sensors.

PM sensor: The Honeywell HPMA115S0-XXX sensor was employed for detecting particle concentrations of $PM_{2.5}$ and PM_{10} in the air. This sensor has an output range of 0 to 1000$\mu g/m^3$ and provides accuracy up to 15%, with a corresponding

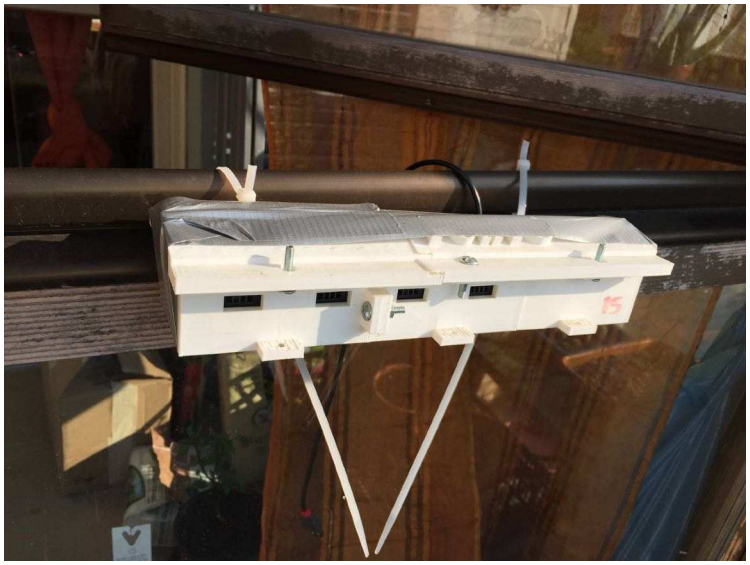


Fig. 2.13 Monitoring Board

price range of approximately $25 to $30 USD. The sensor utilizes the Light Scattering technique, wherein a rotating fan draws air into a chamber, which is then exposed to a laser beam. A proprietary algorithm inside the device estimates particle concentration by analyzing the photodiode response. The sensor communicates with the monitoring board through UART (Universal Asynchronous Receiver Transmitter) communication, writing a new data point every second.

Temperature and humidity: Those measurements are obtained using the DHT22 sensor, which is a low-cost digital sensor known for its affordability and accuracy. It provides humidity readings ranging from 0% to 100% with an accuracy of 2 to 5 percent. Temperature readings range from -40 to 80 with a precision of 0.5 °C, and the sampling rate is approximately 0.5 Hz. Dedicated APIs are utilized to interact with the sensor via One-Wire communication protocol.

Atmospheric pressure: To get these measurements, the BME280 sensor from Bosch is employed. This sensor is known for its precision in measuring barometric pressure and temperature. It has a barometric pressure sensing range of $300-1100\ hPa$, with a resolution of $0.03\ hPa/0.25\ m$. The operational temperature range is 40-85 °C with a temperature precision of 2 degrees Celsius. The sensor data is retrieved through the I2C interface, using specific APIs.

To synchronize the temporal data, a clock for Unix time is utilized. This ensures consistent time references across the measurements. It is important to note that the

experimental data collected by the board stations do not have official calibration and, therefore, do not possess normative value. However, they provide qualitative insights into the system's potential and allow for considerations to be made based on the data obtained.

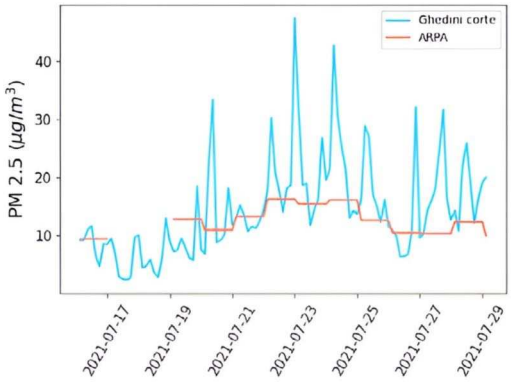
2.2.4 Results Discussion

The results obtained from the experimentation serve two main purposes. Firstly, they aim to assess the advantages of widespread low-cost monitoring compared to the current official monitoring system provided by ARPA. Secondly, they aim to investigate the impact of the urban environment and traffic on the propagation of pollutants. The frequency of monitoring enabled the collection of a substantial dataset for analysis. Throughout the three monitoring campaigns, approximately 30,000 data points were collected for PM levels and approximately 14,000 for temperature and humidity. The collected data were processed again, removing outliers, and grouped into different time frames.

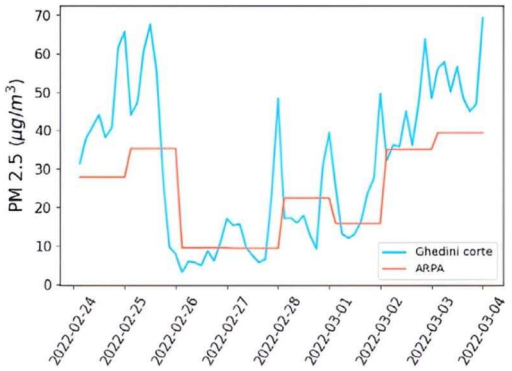
One noteworthy insight is derived from the level of detail in the data. While ARPA only provides a daily average, the low-cost stations can capture hourly and daily trends, thereby revealing any critical periods. The graphs in Figure 2.14 illustrate the $PM_{2.5}$ values from the same sensor during the three monitoring campaigns, compared to the official data from ARPA. Although the trends and daily average values from the low-cost stations align with the official data, it is evident that the daily PM trend can exhibit significant differences. Particularly during the winter period, when PM levels are higher, deviations of up to $25-30\mu g/m^3$ can occur on the same day.

In the first two campaigns conducted during the spring and summer seasons, there is a more pronounced disparity between the monitored data and the official data. This observation is consistent with the non-linearity of the sensors and their lack of accuracy at low concentrations[39]. In other words, the higher the PM value, the greater the margin of accuracy in the collected data.

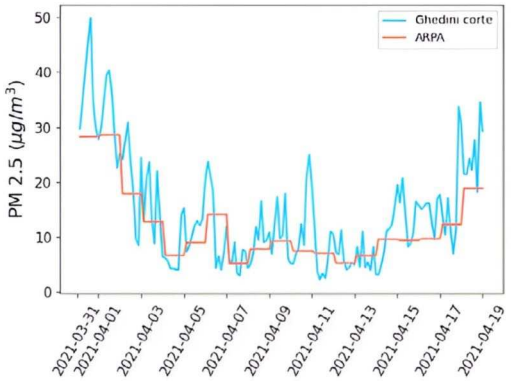
To address the second research question, two primary analyses were conducted. The first analysis compares the data from the four monitoring points during the same period. The second analysis compares a sensor positioned towards the road with another sensor facing the street, which is more exposed to traffic and wind.



(a) $PM_{2.5}$ values in April 2020



(b) $PM_{2.5}$ values in July 2020



(c) $PM_{2.5}$ values in February 2021

Fig. 2.14 $PM_{2.5}$ trends in the three monitoring campaigns (a. April; b. July; c. February). In red, the official data provided by ARPA

Figure 2.15 illustrates the $PM_{2.5}$ trend during the initial monitoring campaign. On average, there is a variation of approximately $6\mu g/m^3$ between the different points, which is relatively small compared to the overall range of values. Throughout all three monitoring campaigns, no significant differences were observed among the sensors at the four monitoring stations.

A more in-depth analysis on the differences between the sensors placed in the inner courtyard (namely "corte" in the figures) and on the balcony facing the road were presented in figure 2.16. By analysing a shorter period, it is possible to highlight the daily trend of $PM2.5$ values without any significant variation. The action of wind and the presence of cars for the specific context considered cannot be assessed. Nevertheless, the simulated wind speed (Figure 2.12) showed homogeneous values in the streets as well in the courtyards of the experimental field, thus confirming possible similar PM values.

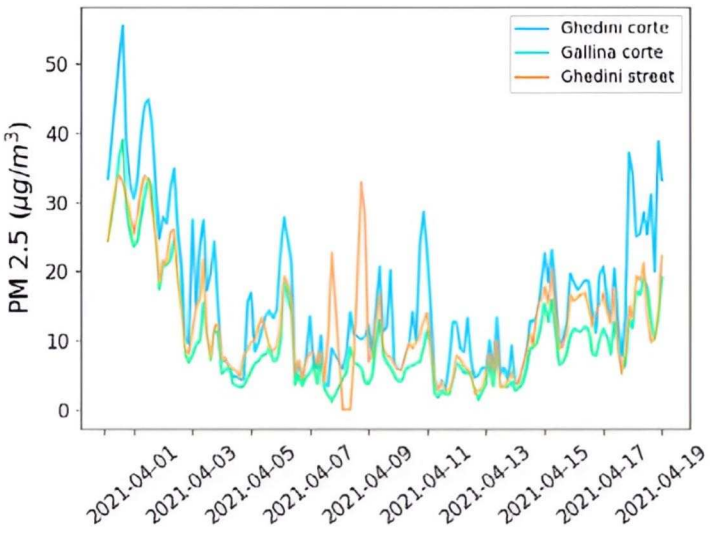


Fig. 2.15 $PM_{2.5}$ values in the same campaign monitoring by different stations.

The findings reveal a concerning air quality situation in Turin. The average values obtained from the monitoring campaigns align with official data, indicating pollutant levels that exceed European limits. Although the proposed technologies and results are still in the experimental stage, they offer potential for monitoring environmental factors that impact neighborhood health. The collected data confirm that the winter season experiences higher pollutant concentrations in the analyzed context.

It should be noted that the qualitative nature of the obtained values stems from the use of sensors that lack official calibration. However, the data calibration provides increased reliability, particularly during the winter period when PM levels are elevated. One limitation of the study is that the analyzed case study did not offer substantial evidence to validate the correlation between the urban fabric and proximity to streets. The limited traffic volume in the area restricts the possibility of conducting more in-depth analysis. Applying the same approach in denser and busier urban areas could yield contrasting outcomes. Additionally, the scale at which air quality is examined is a significant consideration. While air quality is influenced by global and regional factors, local monitoring plays a vital role in comprehending the anthropogenic action and influencing citizen behavior.

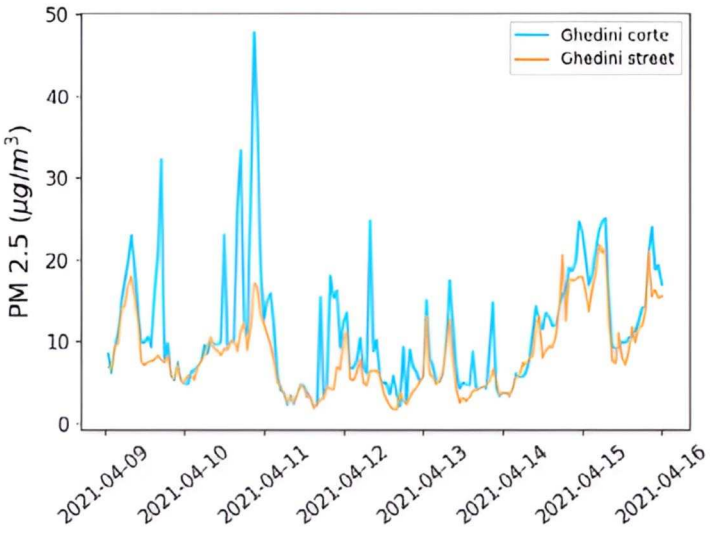


Fig. 2.16 $PM_{2.5}$ values comparison between courtyard (blue) and street (red)

2.3 Conclusions

2.3.1 Indoor Environmental Quality Monitoring

This study emphasizes the significance of Indoor Environmental Quality (IEQ), encompassing thermal, visual, acoustic, and air quality factors, in influencing people's health, comfort, well-being, and productivity. The development of the PROMET&O system is presented, comprising a low-cost and accurate multi-sensor device for ac-

quiring IEQ parameters and a tablet for collecting subjective feedback and providing data visualization to end-users.

The research focuses on the design, development, and metrological characterization of the low-cost multi-sensor device, which incorporates sensors for measuring air temperature, relative humidity, illuminance, sound pressure level, carbon monoxide, carbon dioxide, particulate matter, formaldehyde, and nitrogen dioxide. Calibration procedures based on reference standards ensure the traceability and accuracy of the measurements obtained through PROMET&O.

The verification procedure is conducted to assess the overall measurement chain and confirm whether the measurement errors align with the uncertainty specified by the sensor manufacturer. In cases where the sensor does not meet the uncertainty requirements, a metrological characterization is performed using a reference standard to modify the calibration function of the measurement chain accordingly.

Preliminary experimental results are presented for two specific procedures: the verification of carbon dioxide concentration and the characterization of the illuminance measurement chain, where the uncertainty requirements are not met. The forthcoming measurements with PROMET&O will encompass all monitored parameters to determine whether a verification is sufficient or if a metrological characterization is necessary to ensure measurement traceability for all quantities involved.

The validation of the PROMET&O system will be carried out in a real-world setting, specifically in a company's open-plan office. This validation process will contribute to the advancement of the field by incorporating real-time correlation of objective and subjective data, engaging users through the dashboard and LED visualization of IEQ levels, and promoting energy conservation while enhancing IEQ and IEC.

2.3.2 Outdoor Environmental Quality Monitoring

The last pandemic and the climate emergency showed the need for a significant transformation of urban environments, with a focus on promoting Healthy Cities and aligning with the Sustainable Development Goals [53]. The pandemic has had some positive impacts, such as reduced car traffic leading to improved air quality and perceived road safety, ultimately enhancing the overall quality of life[54].

Moreover, the improvement in air quality and the promotion of pro-environmental citizen behaviors can potentially generate positive externalities, although assessing them can be challenging (e.g., economic benefits from reduced healthcare costs) [54]. Rapid advancements in technology have facilitated a greater understanding of the interconnections between various phenomena in cities [55]. In this regard, ICT and IoT play crucial roles in enabling the ecological transition. Data-driven approaches are increasingly being employed in the management and planning of urban areas, providing a deeper understanding of the factors influencing health and supporting evidence-based decision-making.

Urban planners should rethink and reimagine the urban environment, considering the lessons from history that indicate its evolution after each pandemic. In the context of creating Healthy Cities, the widespread adoption of IoT network for real-time measurements, along with advanced environmental simulations, becomes essential for promoting data-driven design. At the urban neighborhood scale, interventions can have a significant impact, as land-use patterns, building types, technologies, and behaviors that influence environmental quality can vary greatly across the city. Therefore, implementing an auxiliary monitoring network alongside the official one can provide greater awareness of site-specific issues that are closely related to public health.

In today's context, the technological advancements in low-cost sensor systems have reached a level where widespread application in urban settings is feasible. This opens up new possibilities for monitoring and understanding environmental conditions in real-time.

As urban planners, there is a growing responsibility to reimagine urban environments in light of these challenges, as history has shown that cities evolve after each pandemic. In the context of creating Healthy Cities, implementing a widespread IoT network for real-time measurements, coupled with advanced environmental simulations, becomes crucial for promoting data-driven design and decision-making. The urban neighborhood scale is particularly relevant for intervention, as land-use patterns, building types, technologies, and behaviors can significantly impact environmental quality, varying throughout the city. Supplementing the official monitoring network with an auxiliary monitoring system offers greater awareness of site-specific issues directly related to health.

Progress in low-cost sensor technologies, ICT, and data processing algorithms has created highly promising opportunities. However, certain challenges remain: On the one hand, the excessive generation of data can pose difficulties in managing the flow of information and may have environmental implications. On the other hand, issues related to data ownership and privacy still require proper regulation.

Considering the crucial role of data, future research should prioritize the establishment of new public-private partnerships that regulate the production of data in a manner that is both controlled and valuable. This will ensure responsible data management and enable the realization of the full potential offered by these technological advancements.

Chapter 3

Data Collection

Air pollution has emerged as a pressing concern in urban environments, demanding comprehensive strategies for effective monitoring and mitigation. With the rise of smart cities and the Internet of Things (IoT), innovative approaches leveraging low-cost sensor acquisition have gained significant attention. This chapter explores the integration of IoT technologies in air monitoring systems, specifically focusing on the utilization of affordable sensors to measure particulate matter (PM) concentrations.

The first chapter's section delves into the importance of accurately measuring ultra-fine particulate matter (PM) and the development of affordable sensors and wireless sensor networks has facilitated data collection and analysis. However, monitoring ultra-fine particles presents unique challenges, particularly in determining the optimal sampling rate. This chapter presents a comprehensive analysis of the frequency behavior of PM signals using an open-source dataset. The objective is to determine the maximum frequency required for digitizing air pollution signals, enabling the establishment of a sampling rate that ensures accurate estimation of particulate matter concentration while minimizing data and energy consumption.

Moving on the second chapter's section, the focus shifts towards overcoming engineering challenges associated with resource optimization in low-cost light-scattering sensors. While these devices provide valuable data, issues such as energy consumption, data transmission, and sensor aging need to be addressed. This study investigates the possibility of reducing the duty-cycle of an air pollution monitoring sensor while still obtaining meaningful data on its general behavior. By reducing redundant information logging, the strain on the sensor is alleviated, leading to

extended lifespan and reduced operational costs. Additionally, energy consumption is minimized, making it particularly advantageous for battery-powered devices. The study aims to optimize the duty-cycle, ultimately enhancing the efficiency and effectiveness of air pollution monitoring.

By examining the frequency behavior of PM signals and optimizing the duty-cycle of air pollution sensors, this chapter contributes to the advancement of air monitoring in IoT application environments for smart cities. The findings and insights derived from these topics will shed light on the challenges, opportunities, and potential solutions in the quest for cleaner and healthier urban environments.

Some of the work described in this chapter has been previously published in [56], [36].

3.1 Frequency Analysis

As urban areas continue to expand and their populations grow [57], the emission of pollutants has become a significant concern for public health. According to the World Health Organization [58], exposure to air pollutants such as particulate matter (PM_{10} and $PM_{2.5}$) was linked to approximately 4.2 million deaths in 2016. To assess air quality in metropolitan regions and determine contaminant levels, professional and highly accurate monitoring stations are employed (refer to Figure 3.1). However, the cost of each station can reach hundreds of thousands of dollars [59], resulting in most cities having only a limited number of high-precision stations. As a consequence, the spatio-temporal resolution provided by these stations is low since pollutant concentrations can vary even within a small distance, such as tens of meters near an urban intersection [60].

Fig. 3.1 High-precision monitoring station - Rubino station in Turin- Italy

The light-scattering particle (LSP) technique has gained significance in recent years as an affordable method for estimating PM concentration [61]. LSP sensors consist of a light source positioned at a forward angle along with a light sensor. Air is passed through an inlet in front of the light sensor, and when particles scatter the light, the photosensor generates an estimation of the particulate concentration. This technique allows for the development of low-power, compact, and cost-effective sensors (less than US$100), making them ideal for IoT applications. Recent evaluations and comparisons of LSP-based PM sensors have highlighted that the sensors' performance primarily relies on calibration [62]. Environmental factors such as humidity, temperature [63, 64], particle type, and deployment type (mobile, wearable, or fixed) can affect their precision [19]. Therefore, on-site continuous calibration is necessary to maintain accuracy over time. State-of-the-art calibration and network re-calibration models have been examined in studies like [65] and [16]. However, these studies also emphasize that improving accuracy is still an ongoing research area. But LSP sensors, as discussed in section 2.2, demonstrate significant potential for large-scale deployments and urban monitoring.

Numerous projects and prototypes have been developed to enhance the spatio-temporal resolution of PM sensors incorporated in low-cost monitors. European guidelines [66] have started incorporating these sensors to measure air quality, and

test networks for these technologies have received financial support [67]. A comprehensive overview of funded projects can be found in [19]. However, there is currently no established strategy for consolidating the collected data. Crowdsensing and WSN systems have been at the forefront of these efforts [68, 69], but both approaches encounter challenges regarding data acquisition rate and energy consumption. The accuracy of air quality estimation at a specific location depends on factors such as the timing of PM measurements, proximity of monitoring devices, and the communication architecture employed.

Despite the progress made in previous research, it is widely acknowledged that energy optimization remains a crucial area of investigation and improvement in the development of prototypes, particularly for low-cost PM sensors. Energy optimization in air quality monitoring devices can be addressed at three levels: the sensor level, the node level, and the network level [70]. However, the most challenging aspect of optimization lies in comprehending the behavior of pollutant concentrations that need to be sensed. This necessitates signal processing analysis to understand the frequency characteristics of such signals. Future advancements and optimizations at the node and network levels will heavily rely on this understanding of signal behavior in the frequency domain.

Based on the aforementioned considerations, this study conducts a frequency analysis to determine the appropriate width of PM concentration (PM_{10} and $PM_{2.5}$ in $\mu g/m^3$) based on the estimation frequency. The analysis utilizes case studies and a dataset described in [13] to examine and validate various scenarios of PM acquisition and sharp events that indicate rapid changes in PM concentration. Two criteria are employed and compared to ascertain the signal bandwidth and to establish an optimal estimation period through statistical analysis. This approach aims to reduce the data size generated by the PM monitor and optimize resource management. The dataset consists of measurements taken over a five-month period during the autumn-winter season of 2018/2019 in the metropolitan area of Turin, Italy. The dataset is openly accessible and available in the source cited as [71].

3.1.1 Background and Related Work

In response to increasing scientific interest, there has been a surge in the development of air quality monitors in recent years. A significant number of these monitors

incorporate low-cost PM meters. However, there has been limited research conducted to assess the effectiveness of these sensors in capturing abrupt changes in PM concentration, particularly in complex urban environments [60]. The following is a concise overview of the acquisition rates utilized in different studies.

According to the classification proposed by [69], low-cost PM monitor deployments can be categorized into three groups. The first group is the Static Sensor Network (SSN), which involves placing PM monitors at fixed locations of interest, such as intersections or other stationary points. The second group is the Community Sensor Network (CSN), where devices are carried by volunteers who are interested in monitoring air quality. These devices can be attached to personal modes of transportation such as bicycles or electric scooters. The third group is the Vehicle Sensor Network (VSN), which involves deploying devices on specialized vehicles or public transportation. Table 3.1 provides a summary of the most significant projects and their outcomes for each type of low-cost PM monitor deployment in the field.

The current literature trend focuses on prototypes designed for SSN deployments, which require a higher number of sensors to adequately cover a metropolitan area. These implementations have shown the best accuracy results for PM sensors based on LSP technology. On the other hand, CSN deployments offer the advantage of placing sensors in close proximity to users, enabling higher spatio-temporal resolution from the user's perspective. However, the precision of these sensors can be affected by user manipulation errors or rapid movements, such as when the sensor is mounted on a bicycle. In such cases, LSP-based sensors exhibit significant variability in PM estimation, and in a mobile scenario, the airflow inside the sensor becomes random and uncontrolled. Conversely, VSN deployments encounter challenges due to airflow difficulties within low-cost PM sensors [97]. As a result, VSN deployments typically exhibit lower estimation precision [98]. To improve accuracy and mitigate the effects of airflow, some VSN deployments have incorporated machine learning techniques and anemometers.

Regarding the estimation period, the literature analysis provided in Table 3.1 highlights the absence of a defined standard period. Typically, the selection of the estimation period is influenced by factors such as sensor speed (typically around 1 second), resource consumption reduction (usually greater than 1 minute), or the need to reduce sensor noise through time averages. However, the period selection is not directly linked to the frequency behavior of the PM variable. As a result, the

Table 3.1 Summary information of literature more relevant low-cost PM monitors in literature.

Network type	Sensor	Estimation period	Reference
SSN	Shinyei PPD42NS	1 h (avg.)	[72]
		5 min (avg.)	[73]
		30 s	[74]
		10 min	[75]
	Shinyei PPD20V	15 min	[63]
	Shinyei PPD60NS	1 min	[76]
	Sharp GP2Y1010	10 ms - 1 s (avg)	[61], [77]
		10 s	[78]
	Syhitech DSM501A	1 min (avg)	[79]
	Plantower PMS 1003/3003	10 s	[80]
	OPC-N2	5 min (avg)	[81]
	Honeywell HPMA115S0	1 s	[82],[13]
	not defined	1s	[83]
CSN	Sharp GP2Y1010	3 s	[84]
		<1 s	[85],[86]
		1 s (avg)	[87]
		1 min	[59]
	Shinyei PPD42NS	<1 s	[88]
		30 s	[89]
	Shinjei PPD42NJ	5 min	[90]
	Sharp GP2YI051	1 min	[91]
	Sharp DN7C3CA006	1 min (avg)	[92]
	Samyoung DSM501A	<1 s	[86]
	Dylos DC1700	1 min	[93]
VSN	unknown Sharp model	5 s	[94]
	Shinyei PPD42NS	1 s	[95]
	MiniDiSC	5 s (avg)	[96]

estimation process may involve capturing redundant and inefficient measurements. This approach can potentially overlook specific events, such as rapid changes in PM concentration during peak hours near a vehicular crossing.

In [99], an analysis is conducted to reduce the duty cycle of PM sensors and, consequently, lower energy consumption. However, this analysis is solely based on data distribution and does not incorporate frequency analysis. On the other hand, [100] provides a comprehensive examination of the frequency behavior of various pollutants, including PM_{10}. The study utilizes measurements from professional stations, which have a maximum estimation rate of one hour. Unfortunately, this rate is insufficient for analyzing short-duration events with variable concentrations that can occur in urban environments, thereby limiting the spatio-temporal resolution.

The estimation period for low-cost sensors, including popular models like the Sharp GP2Y1010 and Shinyei PPD42NS, varies significantly across different studies, with no consistent pattern observed. In certain cases such as [85, 88], and [86], the maximum sampling speed specified by the sensor manufacturer or to the detection of sudden atmospheric phenomena, as described in [13], are used as reference points for setting the estimation period. On the other hand, some studies adjust the period based on measurements from reference equipment, as demonstrated in [61] and [95].

3.1.2 Data and Methodology

Data set description

The dataset consists of measurements from multiple monitors, each equipped with six sensors: four Honeywell HPMA115S0-XXX sensors for estimating PM_{10} and $PM_{2.5}$ levels, a DHT22 sensor for temperature and relative humidity measurements, and a Bosch BME280 sensor for atmospheric pressure. The monitoring system is controlled by a Raspberry Pi Zero Wireless, which features a 1 GHz ARM 11 microprocessor and provides 2.4 GHz Wi-Fi (802.11 b/g/n) and Bluetooth 4.1 connectivity. A block diagram of the pollution monitor is depicted in Figure 3.2.

The dataset comprises data from various experiments involving stationary measurements conducted by twelve monitors located near the reference air quality measurement station in Rubino, located in the city of Turin, Italy (refer to Figure 3.3). The data was collected from October 2018 to January 2019, covering the autumn and

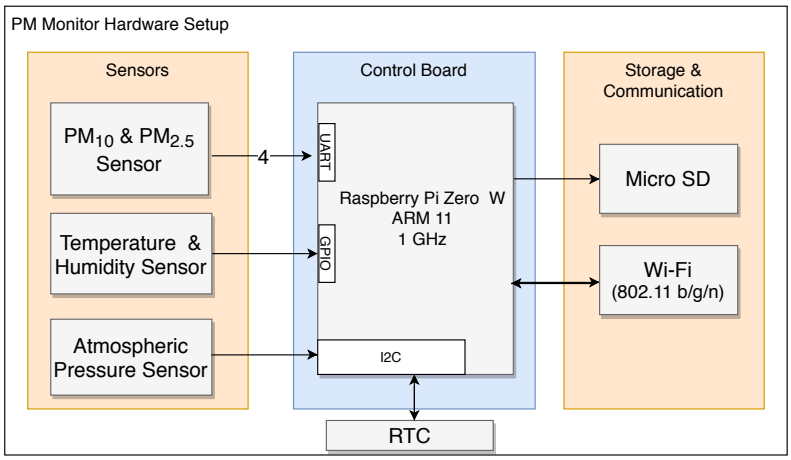


Fig. 3.2 The monitor device's hardware block diagram, it includes six sensors, a micro SD card, and a Wi-Fi connection for data upload to an external server.

winter seasons known for higher pollution levels, as reported in [101]. Additionally, the dataset includes mobile measurements captured during dynamic applications to investigate specific instances of atmospheric phenomena. It is important to note that the data collected from PM sensors was acquired at the maximum sampling rate specified by the manufacturer [102] to capture higher-frequency concentration variations. Temperature, humidity, and pressure measurements were not included in the Fourier analysis.

(a). (b).

Fig. 3.3 Sensor Deployment on Rubino Station (**a**) Low-cost monitor arrangement for data collection. (**b**) Deployment of monitors on the Rubino station roof

Post-processing tasks

In the post-processing stage, Figure 3.4 provides an overview of the data processing workflow. Initially, the RAW data is retrieved from the *MySQL* database and compiled into a RAW .csv file. Since the hardware platform relies on an operating system (OS), a pre-analysis sub-process is conducted to detect errors caused by hardware malfunctions, measurement process crashes, or temporal inconsistencies. The latter issue arises from the system's reliance on a microprocessor and a non-real-time OS, which can result in irregular acquisition times and periods of missing data. The data then undergo a Data Correction sub-process, where values lacking coherence or exhibiting inconsistencies are removed. In cases where data gaps of 10 seconds or less are identified, a linear interpolation is applied to fill in the missing values. This interpolation introduces minimal low-frequency noise that has a minimal impact on the frequency analysis results in subsequent steps.

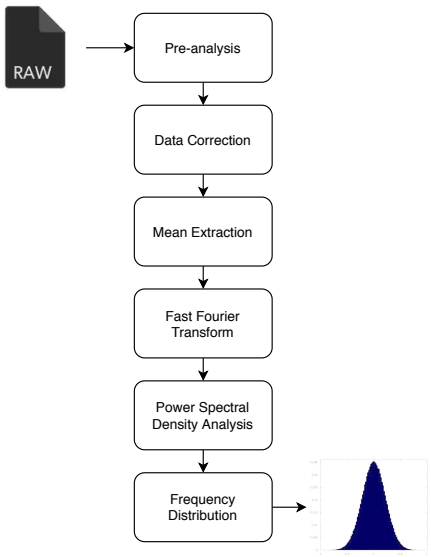


Fig. 3.4 Post-Processing Overview

The subsequent stages of the process follow the methodology outlined in [103]. In these steps, the mean value is extracted (*Mean Extraction sub-process*) to remove the DC component from the signal. Afterward, the *Fast Fourier Transform* (FFT) is applied to the data series within a daily window containing more than 21,600 samples. This window is selected to ensure that the data does not have prolonged periods of missing values or falls within a maintenance window, as these events can introduce windowing noise.

Finally, the *Power Spectral Density Analysis* is obtained to determine the bandwidth of the signal, which represents the maximum frequency range containing the data series. Estimating the bandwidth of an unknown signal is a complex task and often relies on empirical methods. Different definitions, such as the one in [104], define the bandwidth as the spectrum portion that accumulates 95% of the energy. Other approaches involve statistical definitions, such as the empirical rule or confidence intervals. In order to comprehensively analyze the frequency trade-off, both percentage criteria (95-99%) are evaluated based on the Power Spectral Density (PSD) of the data. This analysis yields the *Frequency Distribution* for each particulate material sensor, obtained from the entire duration of the experiments.

Experiment analysis

In this last subsection, we examine sharp pollution events that were measured in [13]. These events encompass pollution measurements near a traffic light, a wind event, and New Year's Eve fireworks. However, the citizens' mobility study presented in [13] could introduce uncertainties in airflow and result in imprecise measurements as was discussed in 3.1.1. Therefore, these experiments are not considered in the analysis presented in this research.

3.1.3 Results and Discussion

Bandwidth criteria

Prior to the analysis of the data set, an initial iteration was conducted to gain insights into the behavior of the PM sensors. A short data set with minimal PM variability was utilized to assess the sensor's internal noise. The power spectral density (PSD) was computed and visualized as a histogram (refer to Figure 3.5). The histogram exhibited two distinct Gaussian distributions: one near $-100dB$ and another peak close to $-60dB$. The latter peak corresponded to the signal-to-noise ratio (SNR) of a 10-bit analog-to-digital converter (ADC), aligning with the sensing range of the PM sensor, which is defined in 10 bits. The $-100dB$ peak represented the floor noise, indicating that utilizing 99% of the PSD to determine the bandwidth would not yield any substantial gain as most of the bandwidth would consist of noise attempting to reach the energy criteria. Conversely, the empirical law defines a confidence interval

of $\mu + 2\sigma$, representing approximately 95% confidence, with an expected frequency event of 1 in 22 events falling outside of the confidence interval. Thus, choosing the 95% criteria strikes a suitable balance between estimation error and accurately representing the frequency of PM pollution events.

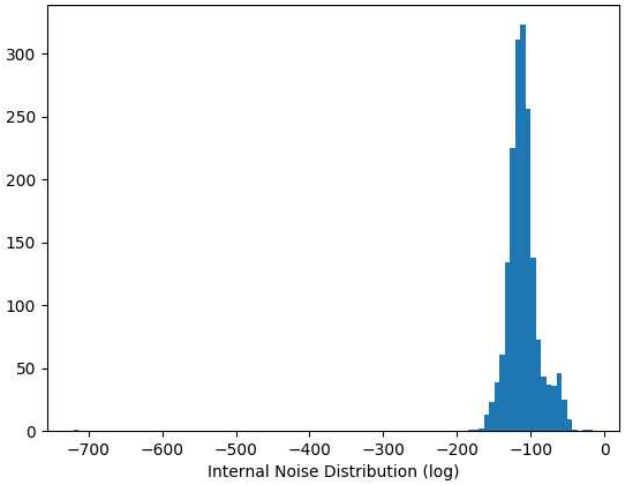


Fig. 3.5 Sensor internal noise PSD histogram

3.1.4 Stationary frequency analysis

The analysis was conducted for each individual monitor, and the distributions of bandwidth for PM_{10} and $PM_{2.5}$ were derived from the acquired data. However, three monitors (with board IDs 4, 7, and 9) exhibited inconsistent values in their data sets, rendering their results unsuitable for the study and thus excluded. Additionally, sensor 2 of monitor ID 2 experienced a failure, producing measurements with concentrations of 0 or 1 for both PM_{10} and $PM_{2.5}$. Consequently, only the data obtained before the failure occurred was considered for the analysis. For the resulting data, the bandwidth was calculated in daily windows to indicate the maximum variability of concentration observed by a sensor throughout the day. Figure 3.6 illustrates the bandwidth distribution results for monitors with ID 5 and 8, which had the highest number of data points collected during the experiment.

From Figure 3.6, it can be observed that the discrete distribution indicates that approximately $\sim 70\%$ of the time, the PM bandwidth is below $0.01Hz$. These

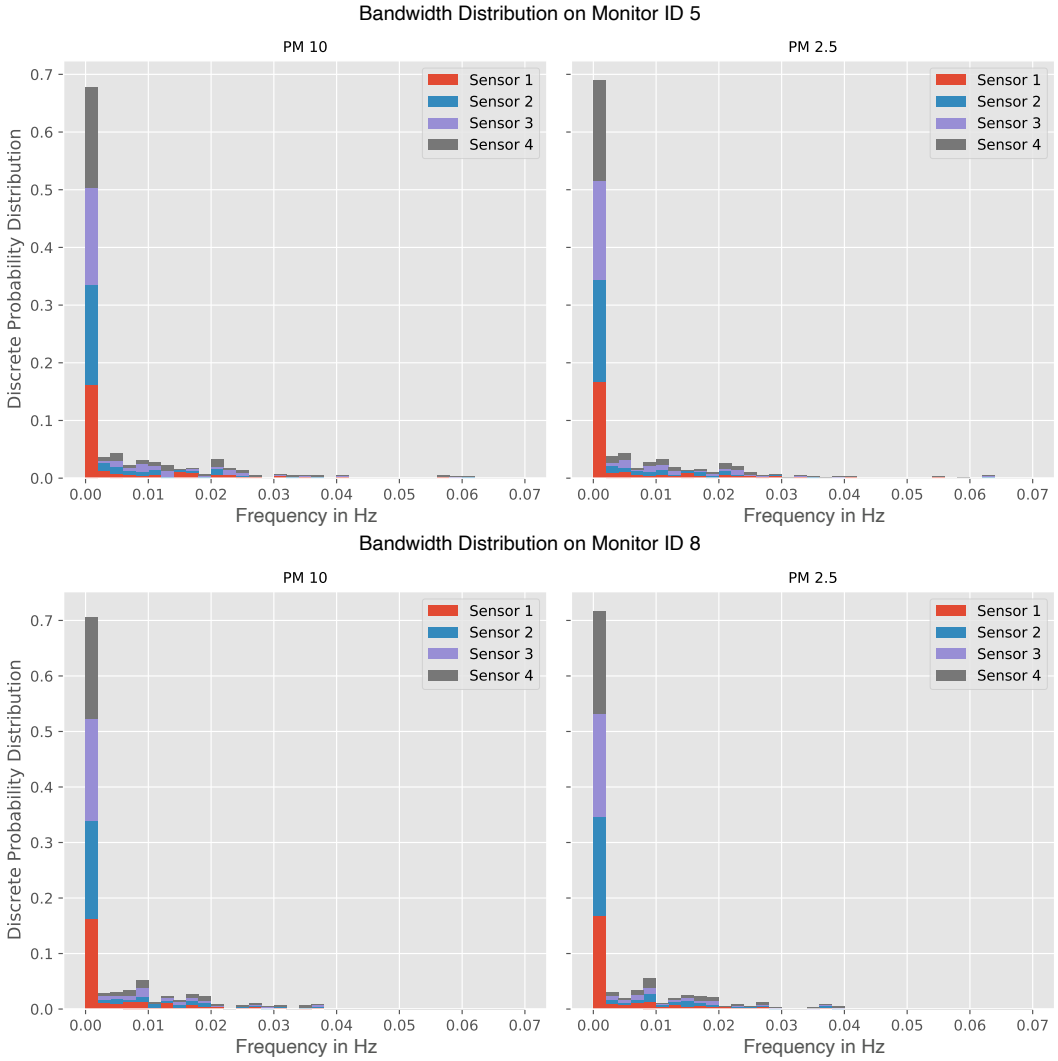


Fig. 3.6 PM_{10} and $PM_{2.5}$ discrete probability distribution of PM bandwidth for monitor ID 5 (*top*), 8 (*bottom*) for 95% of the PSD.

Table 3.2 General bandwidth statistics for each monitor in Hz.

Monitor ID	Mean $PM_{2.5}$	Mean PM_{10}	σ $PM_{2.5}$	σ PM_{10}	Maximum $PM_{2.5}$	Maximum PM_{10}
0	0.01007	0.01065	0.01309	0.01337	0.04296	0.04171
1	0.00880	0.00923	0.01172	0.01198	0.04263	0.04155
2	0.00418	0.00436	0.00805	0.00825	0.04193	0.04169
3	0.00557	0.00573	0.00870	0.00891	0.04166	0.04089
5	0.00496	0.00514	0.00941	0.00955	0.06347	0.06143
6	0.00495	0.00516	0.00849	0.00870	0.03941	0.03884
8	0.00408	0.00424	0.00747	0.00766	0.03853	0.03715
10	0.00519	0.00536	0.00865	0.00888	0.04011	0.04070
11	0.00501	0.00523	0.00889	0.00916	0.04234	0.04303

findings align with the results reported in [100] for PM_{10} pollutants. Consequently, fast estimation periods have minimal impact on capturing PM values. The mean value of the data exhibits a bandwidth of approximately 0.006 Hz throughout the entire duration of the experiment. Therefore, a reduced PM estimation period can be determined based on the bandwidth requirements. However, selecting an appropriate estimation or sampling period is not a straightforward task as it should accommodate all atmospheric phenomena present in an urban environment. Table 3.2 summarizes the mean and variation of the PM signal bandwidth obtained for each monitor from the stationary dataset measurements.

The information presented in Table 3.2 reveals that the mean bandwidth value is approximately 0.006 Hz, with an average standard deviation of around 0.009 Hz. This suggests that a significant portion of the particulate material signal's behavior occurs at frequencies below 0.015 Hz. According to the Nyquist-Shannon sampling theorem, an estimation period of around 30 seconds would be sufficient for the PM sensor to capture most common pollution events in an urban environment. However, it's important to note that less frequent sharp PM events may go undetected at this estimation period. Therefore, considering the maximum frequency values, an alternative estimation period of approximately 10 seconds per measurement could be considered. Nevertheless, it's worth mentioning that this would require significant computational resources and could pose challenges for large-scale deployments.

Experiments' bandwidth analysis

To gain insights into sharp events, a similar frequency analysis was conducted using data from the study by [13]. On October 28th, a föhn wind event occurred in Turin, resulting in a decrease in PM concentration levels. In Figure 3.7, the PSD plot is presented, and the red line represents the bandwidth limit on that particular date. It is worth noting that this event is challenging to detect in the frequency domain due to its slower frequency components compared to the faster estimation rate. In contrast, the fireworks event shown in Figure 3.8 and the traffic light measurements in Figure 3.9 exhibit frequency components that surpass the internal sensor noise level. This indicates that if the estimation period is set too close to twice the bandwidth, some information may be missed. However, it should be observed that shorter estimation periods yield negligible contributions since their occurrence probability is minimal, as demonstrated in Figure 3.6. Selecting an optimal estimation period reduces the amount of data to be transmitted, enables the use of narrowband communication technologies, and facilitates the application of duty cycles to the sensor for energy consumption reduction.

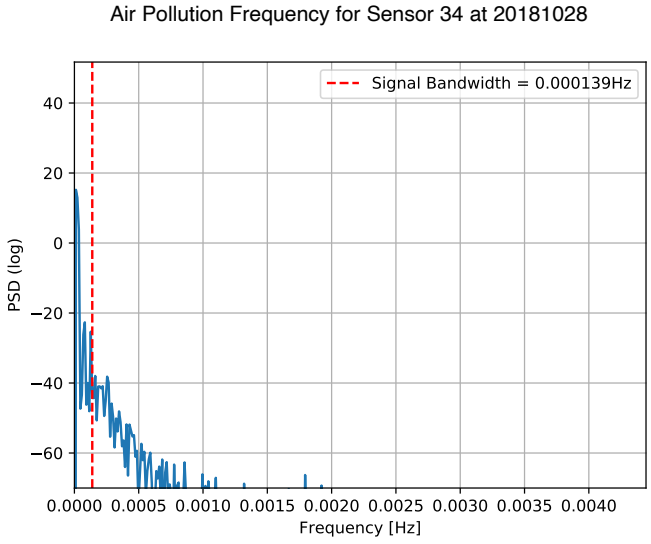


Fig. 3.7 Power spectral density during the föhn wind event for $PM_{2.5}$sensor

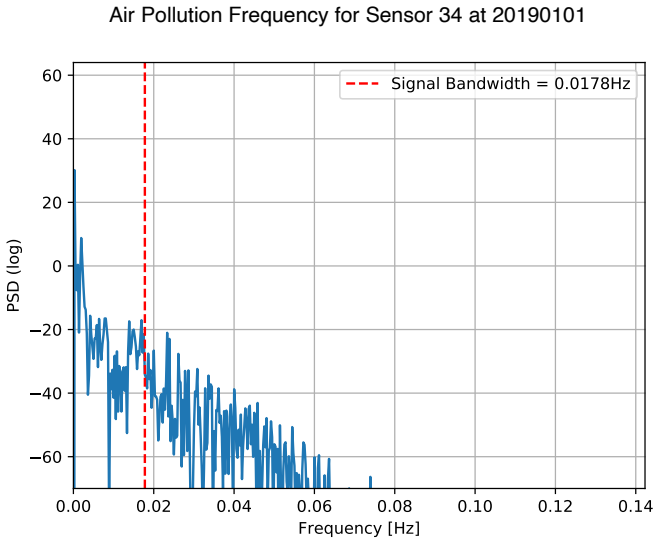


Fig. 3.8 Power spectral density during New Year's Eve fireworks event for $PM_{2.5}$sensor

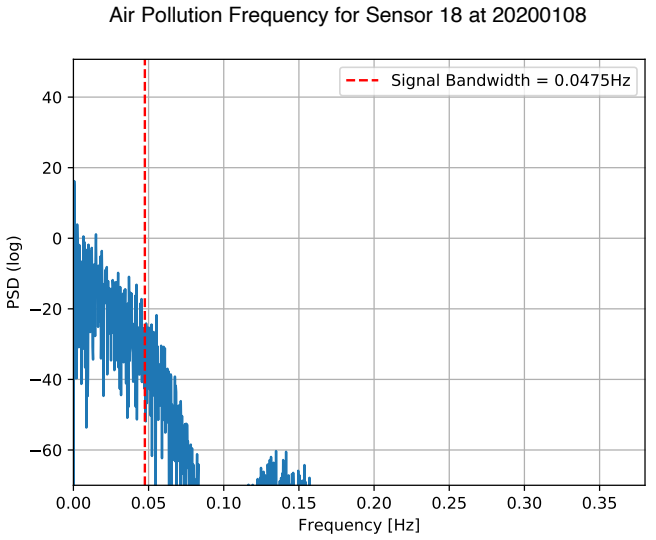


Fig. 3.9 Power spectral density during traffic event for $PM_{2.5}$sensor

Energy Consumption

Considering the lower estimation frequency obtained from the previous results, it becomes feasible to define a duty cycle that reduces energy consumption. The sensors utilized in this study have an approximate current consumption of $80mA$ during active periods and $20mA$ in standby mode (at $5V$). Based on the results, an estimation frequency of approximately $30s$ is sufficient. Therefore, a duty cycle of 20% (6 seconds) within a cycle period of $30s$ could be effective. In this scenario, the energy consumption for this sensor would be only $4.8Joules$. Comparing this value to the consumption of the current prototype, which employs a one-second estimation period (24 Joules of energy consumption per minute), the energy reduction would be approximately 60% (refer to Figure 3.10, left) within the same period.

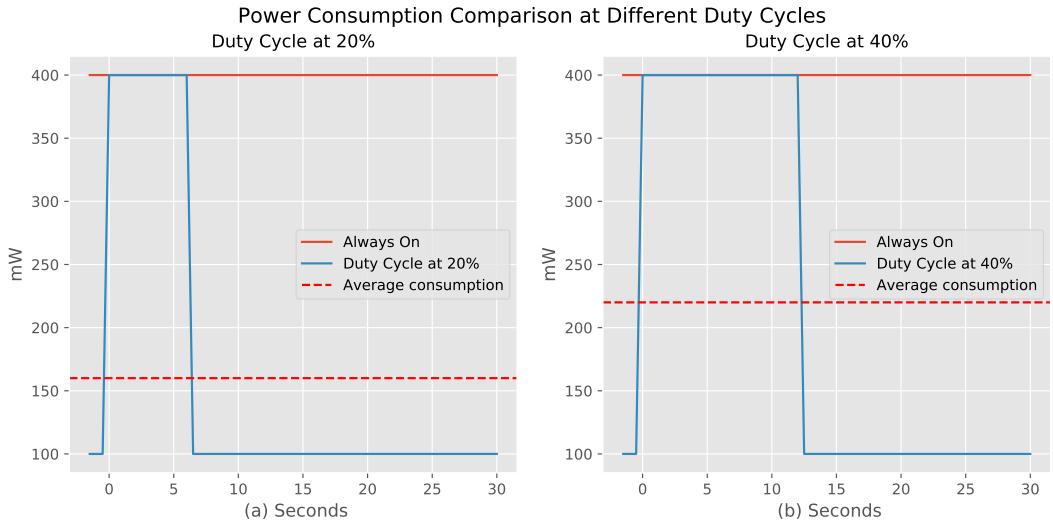


Fig. 3.10 Power consumption comparison for a Honeywell HPMA115S0-XXX sensor for 1 minute with different duty cycles. (**a**)Shows the power consumption at 20% of duty cycle, equivalent to 0.16 Watts (**b**)Shows the power consumption at 40% of duty cycle, equivalent to 0.22 Watts

However, due to the operating environment and limitations of the sensors under LSP technology, a short duty cycle may not guarantee the response times specified by the manufacturer. Moreover, it can lead to the accumulation of particulate material inside the sensor, resulting in erroneous measurements upon activation. To mitigate this issue, a longer duty cycle would allow the airflow to reestablish inside the sensor, thereby avoiding PM concentration estimation errors. In this case, considering a cycle period of $30s$ and a duty cycle of 40% (12 seconds), the energy consumption

would be only 6.6*Joules*. A 40% duty cycle is selected since on average PM sensors under LSP have response times of about 10 seconds. This gives a tolerance value that allows stabilization of the air flow in the sensor. This represents a reduction of 45% in energy consumption per duty cycle of operation (refer to Figure 3.10, right).

The low-cost PM sensors based on LSP technology have demonstrated their utility by detecting variations that professional sensors may not be able to capture. However, it is important to acknowledge that these sensors perform best in stable conditions [39]. In rapidly changing environments or situations where there are variations in the sensor's air intake, the concentration estimates may be inaccurate. Such cases are more common in mobile environments. Nevertheless, the low variability observed among the different sensors and monitors provides a reasonable level of confidence in the reliability of the obtained data.

3.2 Data Gathering optimization

As it was presented in section 3.1, there has been a growing focus within the scientific community on the health impacts associated with air pollution, with various factors contributing to this issue such as the burning of fossil fuels, industrial activities, and wildfires[105]. Among the most detrimental pollutants are Ultrafine Particles, commonly referred to as PM_{10} and $PM_{2.5}$, with diameters smaller than $10\mu m$ and $2.5\mu m$ respectively. Currently, European municipalities typically rely on a limited number of highly accurate sensors as part of their regulatory efforts governed by regulations 50/2008/CE and 107/2004/CE [106].

Regrettably, the use of such precise sensors for creating detailed air pollution maps is hindered by their high cost, large size, and specific operational requirements. These reference-grade sensors can amount to approximately $50,000 - 100,000$ USD in terms of cost[107], weigh around 20 kg, and necessitate high-voltage power supplies. This limitation is evident in the case of Turin, Italy, where ARPA operates 20 monitoring stations to cover an area of approximately 7000 km^2[108] as shown in Figure. 3.11.

Conversely, there is a growing trend of low-cost portable PM monitoring devices that utilize LSP technology. These sensors operate by drawing air into a detection chamber using a fan or a heater resistor. Within the chamber, a low-power light

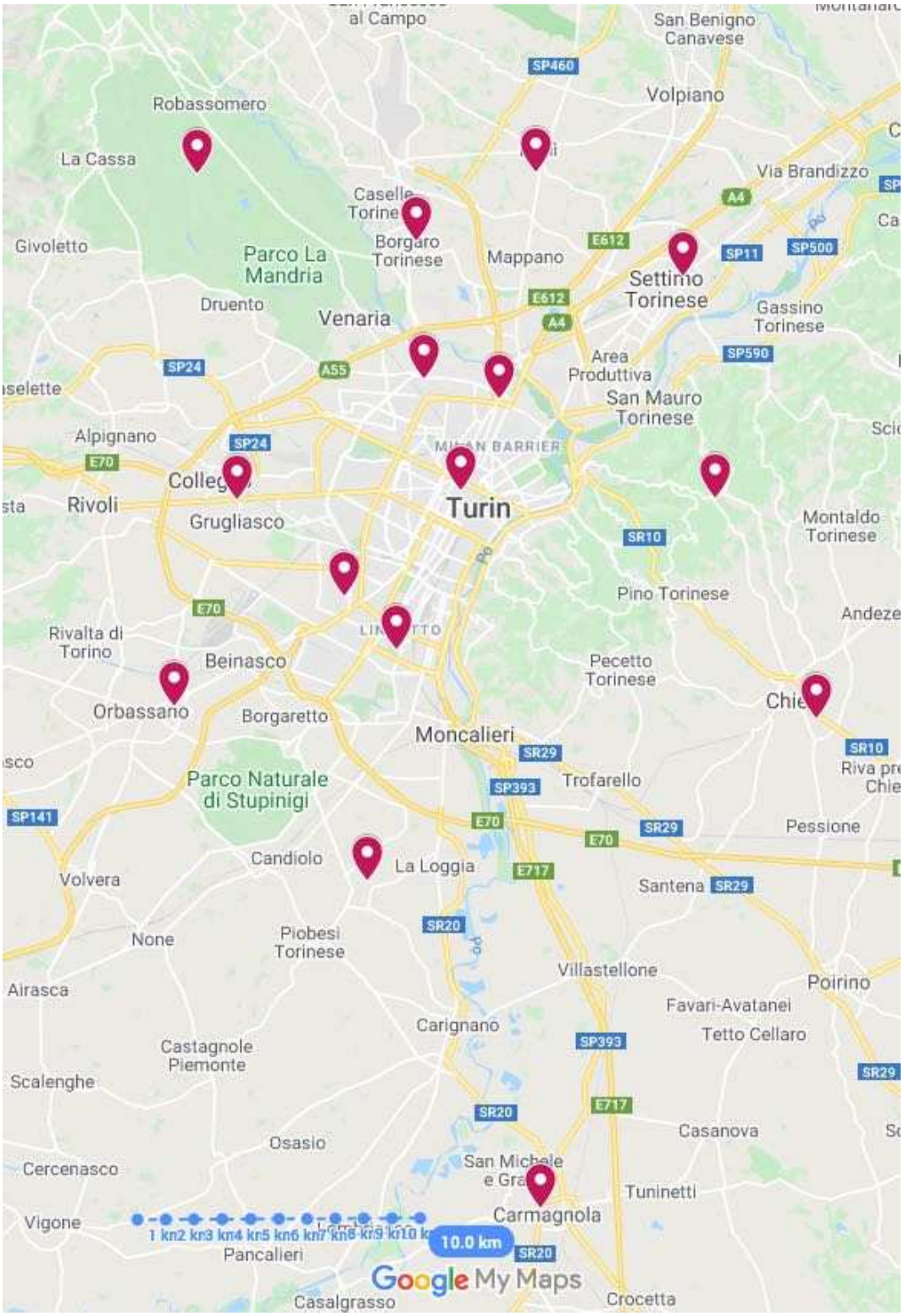


Fig. 3.11 Map of ARPA monitoring stations in the Metropolitan Area of Turin, Italy.

source, typically a laser, illuminates the airflow. The particles present in the air cause the light to scatter, which is then detected by a photo-diode. These sensors are available at a wide range of prices, typically ranging from 30 to $2,000$ USD, and their accuracy may vary[62].

The concentration of air pollution exhibits significant spatial variability, influenced by various factors such as the presence of pollution sources and the configuration of buildings in a specific area[60]. This variation highlights the need for more detailed air pollution monitoring initiatives. To achieve this, the use of affordable and portable devices that maintain reliability can help reduce installation costs while expanding the reach and extent of the monitoring network.

In the context of air pollution monitoring, the utilization of IoT offers an alternative solution to sense and provide accurate PM concentration maps. However, the design of IoT hardware devices must address several important constraints. These constraints include considerations related to network availability, power limitations, and data reliability.

LSP sensors, commonly used in IoT devices, are capable of providing PM concentration values at intervals of approximately 1 second. However, continuous measurements at such short intervals can significantly increase the volume of data that needs to be processed and transmitted. Additionally, the power consumption of these sensors is mainly attributed to the fan or heater resistor incorporated within the device. In the case of sensors used for measuring PM_{10} and $PM_{2.5}$, their current consumption typically ranges around 90 mA [109]. This consumption level must be taken into account, especially in deployments with limited network throughput or restricted power budgets for data transmission. Balancing the need for frequent measurements with the available network and power resources becomes crucial in order to ensure optimal battery autonomy.

Various energy management schemes and techniques have been explored in the literature to extend the battery autonomy of sensors. In the case of power-hungry sensors like LSP sensors, several methods have been proposed, including data reduction, data prediction, and sleep/wake-up (duty cycle) techniques. These methods aim to optimize energy usage and prolong battery life. Among these techniques, the duty cycle method has been found to be particularly effective in saving energy for LSP sensors. By implementing a duty cycle, the sensor can reduce its power consumption to as low as $20mA$ by shutting off the fan or the heater during

the sleep period. This feature is especially desirable for mobile or battery-dependent devices where energy efficiency is crucial. Studies such as Ejaz et al. [110] and Engmann et al. [111] have explored and discussed the benefits of these energy-saving techniques for LSP sensors.

However, LSP sensors encounter challenges in achieving accurate data readings [19]. One of the main challenges arises from the phenomenon of light scattering, which is influenced by various factors such as the composition and shape of the particles, as well as environmental conditions including temperature and relative humidity. To address this challenge, some experiments have demonstrated improved precision performance through in-situ calibration methods [62].

Another source of error in LSP sensors is the aging and drift caused by the accumulation of dust particles inside the air chamber. This buildup can hinder the intensity of light reaching the photo-diode or decrease the overall light power [112]. Therefore, when the fan is shut off as part of the energy-saving measures, it can lead to reduced airflow inside the detection chamber, potentially exacerbating the accumulation of particulate matter. This, in turn, can have a negative impact on the accuracy of the measurements.

The study presented in this subsection investigates the impact of accuracy drift and error in LSP sensors under different duty cycles in a real-world scenario for PM measurements. The goal is to assess the repeatability of measurements and determine the extent of measurement degradation when compared to an always-on LSP sensor.

3.2.1 Background

The literature on low-cost air pollution monitoring is extensive [13, 113–115, 97, 95, 116, 99, 117], with various studies exploring the measurement of air pollution in indoor and outdoor environments. Several papers examine the performance of different sampling frequencies and their relationship to energy consumption. Some of these studies aim to optimize energy usage by reducing power consumption while considering its impact on sensor accuracy. Certain articles discuss different sampling periods or the implementation of duty cycles to extend battery life. However, to the best of our knowledge, no paper comprehensively investigates the effect of long intervals between subsequent samples on measurement accuracy. While some papers

mention this as a potential area for future research, it remains relatively unexplored in the existing literature.

In the paper by Jelicic et al. [115], the authors propose a power management technique to schedule the operation of energy-intensive sensors and wireless transceivers. Their approach aims to reduce energy consumption through three main contributions. Firstly, at the sensor level, they employ a pulse-like mode for the gas sensor and implement early detection of safe concentration conditions. Secondly, at the node level, they manage deep sleep and duty-cycled activity based on the presence of people in the area. Lastly, at the network level, they focus on enhancing the lifetime of each node and leveraging information from neighboring nodes. However, it is worth noting that this paper does not include any analysis of the sampling rate used for sensor operation or the energy efficiency achieved through their proposed techniques.

The authors of [97] present a straightforward air quality monitoring system based on a wireless sensor network (WSN) designed for industrial and urban areas. They propose an energy-efficient routing protocol called CPAS (Clustering Protocol of Air Sensor network) where multiple base stations periodically exchange data. However, the paper does not offer any specific rationale for the chosen sampling rate and selects a sensing period ranging from 200 to 300 seconds without providing justification for this interval.

The paper [95] introduces Mosaic, a mobile sensing system designed for low-cost urban air quality monitoring. The system incorporates a $PM_{2.5}$ sensor and an anemometer to consider air flow velocity when deployed on public transportation vehicles. The authors provide details about the design, implementation, and evaluation of the system, highlighting a novel algorithm for selecting buses to achieve extensive coverage with minimal computational requirements. However, the paper lacks a clear description of the system's energy autonomy and the precision of the model used.

The authors of [116] conducted research on the utilization of wireless sensor networks (WSN) for air pollution monitoring projects in Mauritius. They developed an algorithm aimed at aggregating and merging data, employing filtering and summarization techniques to reduce transmission costs. However, it is important to consider that in a crowd sensing application, the availability of data relies on the data gathering rate. Consequently, a real-time pollution map may not be able to display the current

pollution estimation, particularly when utilizing mobile nodes over multi-hop WSNs, as data will only be accessible when a sensing device enters the range of a collector node. Furthermore, this paper does not include any measurements or evaluations to assess the energy efficiency it aims to achieve.

In [117], the authors introduce a data reduction technique that involves dynamic subsampling of the measured variable, data fusion from multiple sensors for the same variable, and data scaling based on variable ranges. These enhancements aim to decrease the power consumption of the device during both data sensing and transmission. According to their claims, they are able to reduce data collection to only 4% of the raw data without significantly affecting the temporal variability of the pollution concentration. However, the description of their noise model is not clear. Furthermore, they solely utilize raw data from a single LSP sensor without comparing those measurements to other instances of the same sensor.

In [99], the authors put forth a duty-cycling approach for an air pollution monitoring system. This approach focuses on minimizing energy consumption by utilizing a machine learning algorithm that predicts values during sensor inactivity periods. The authors present promising outcomes regarding energy reduction. However, they do not provide details regarding the concentration variations in the test scenarios and emphasize that this model necessitates continuous training. Consequently, it remains unclear whether this process requires keeping the sensors active or if the system is resilient to interruptions.

The objective of this study is to analyze the influence of a duty-cycling approach in urban settings, considering various estimation periods for particulate matter measurements within a resilient scheme. This scheme ensures accurate readings even in cases of temporary sensor failure, such as material obstruction within the sensing chamber. Additionally, in the previous section 3.1(published in [56]), a frequency analysis is conducted to examine the frequency characteristics of ultra-fine particles and suggest different sampling periods for PM measurements in diverse urban scenarios.

3.2.2 Experimental Setup

The purpose of this experiment was to assess the impact of implementing a duty-cycle on sensor accuracy by measuring PM_{10} and $PM_{2.5}$ values in an urban environment.

This section provides details on the selection of PM sensors, the setup of hardware and software components, and the execution of the data acquisition campaign.

Sensor Selection

There are various low-cost LSP sensor models available, and some of them feature a heater resistor that requires a specific orientation, typically in a vertical position, to ensure proper airflow. This orientation constraint adds complexity to the design of an enclosure, especially when integrating such a sensor into a mobile device. Another limitation of sensors with heater resistors is the stabilization time, which refers to the time required to heat up the resistor, establish the airflow, and provide accurate measurements. This stabilization time typically ranges from one to two minutes. As a result, applying a duty cycle to these sensors would be feasible only for measurement periods longer than two minutes.

In contrast, fan-based LSP sensors offer faster stabilization times, typically around 10 seconds, and are more tolerant of changes in orientation. These sensors also provide digital outputs, which simplify the integration process by reducing the need for additional electronics to adapt the signal for different processing units.

Considering the factors discussed, the *HPMA115S0-XXX* sensor [102] was selected for the experiment. This sensor is a fan-based LSP sensor that offers even lower stabilization times (less than 6 seconds) and communicates through a *UART* interface, allowing for the retrieval of measurement information and control of the fan state. The chosen sensor has been demonstrated to exhibit high consistency and coherence in previous studies [82], making it suitable for use as a power benchmark in conjunction with the prototype described in [13]. It has a concentration sensing range of 0 $\mu g/m^3$ to 1,000 $\mu g/m^3$ for both PM_{10} and $PM_{2.5}$ values. The sensor offers two measurement modes:

- **Auto-send mode (default):** where PM measurements are automatically delivered every second
- **Query mode:** where PM measurements can be requested through a query message.

Hardware architecture

The hardware setup described in this section represents an enhanced version of the prototype introduced in [13] and analized in the previous section 3.1. The initial prototype utilized a Raspberry Pi Zero W, featuring a microprocessor capable of operating at 1 GHz and running a Linux operating system to handle measurement and synchronization tasks. However, this microprocessor lacked energy management options to optimize power consumption, even under low computational loads.

To address those limitations, a new monitoring station was designed. The main board was replaced with a Pycom FiPy board [118], which integrates an Expressif ESP32 [119] microcontroller unit (MCU) as the central processing unit. The ESP32 MCU offers various energy-saving modes to reduce power consumption, including an ultra-low-power mode that can achieve consumption as low as $25\mu A$. In terms of communication capabilities, the ESP32 MCU includes built-in WiFi (802.11 b/g/n at 2.4 GHz) and Bluetooth (v4.2 and BLE) connectivity.

Additionally, the FiPy board incorporates Sigfox/LoRA and LTE-M (CAT-M1 and NB-IoT) radio modules, expanding the device's communication options. With these five communication technologies, the device becomes versatile and well-suited for various applications in the context of IoT.

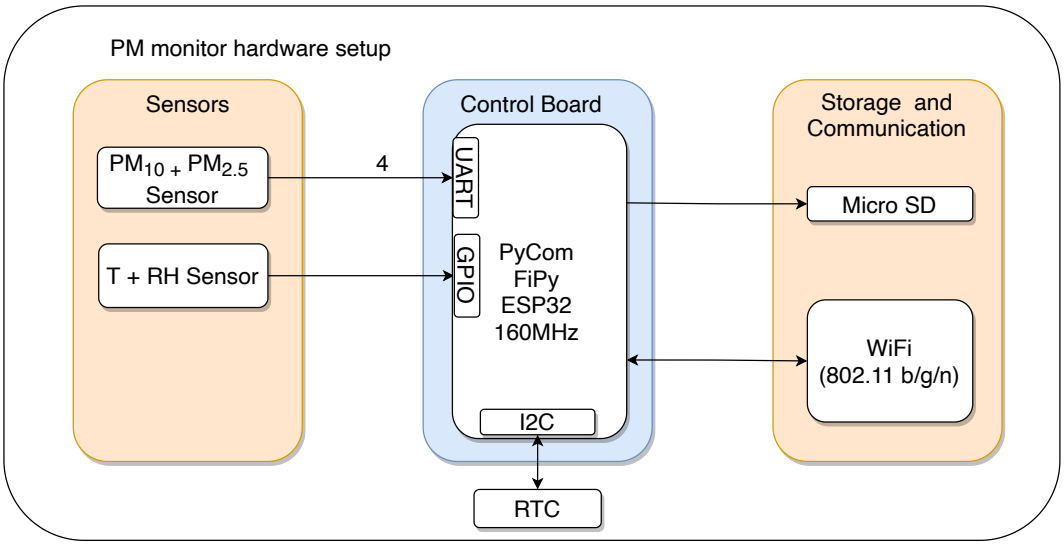


Fig. 3.12 Hardware architecture.

The hardware architecture is illustrated in Figure 3.12, and it includes several peripherals for sensing and recording relevant data related to particulate matter. These peripherals are as follows:

- Four *HPMA115S0-XXX* PM sensors connected to the UART interface of the Fipy board. The FiPy board has two UART ports that can be assigned to any available GPIO pins. One of the ports is used for programming purposes, while the second port is utilized for bidirectional communication with the PM sensors. The software controls the UART assignment to manage and measure each Honeywell PM device (refer to Fig. 3.13).
- One *DTH22* sensor connected via a one-wire protocol to measure temperature and relative humidity.
- One *BMP280* sensor connected through I2C protocol to obtain the atmospheric pressure. But, this sensor is not employed during this chapter.
- One *P1010D* GPS connected through I2C protocol to determine the monitor's geo-referenced position. As the BMP280 this sensor is not used or analyzed in this chapter.
- An external *DS3231* RTC (Real-Time Clock) module connected through the I2C protocol to obtain the current time. The RTC's date and time are updated using the NTP protocol during booting if a known WiFi connection is available.

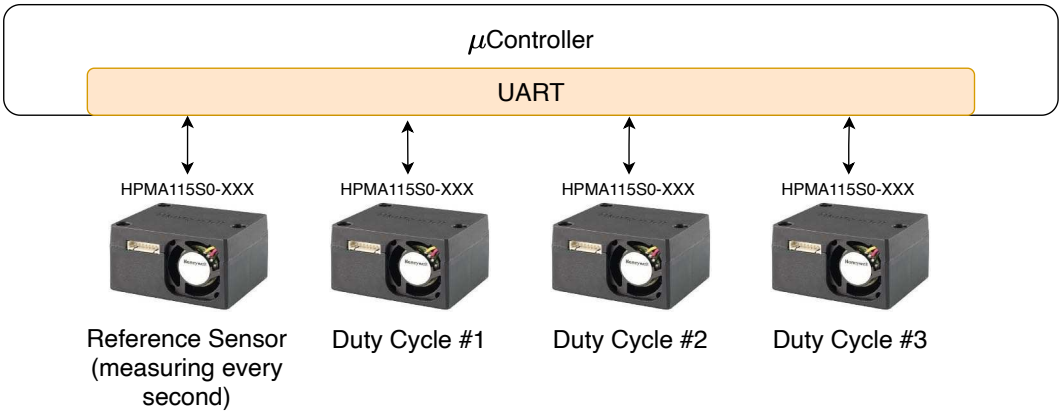


Fig. 3.13 Experiment setup.

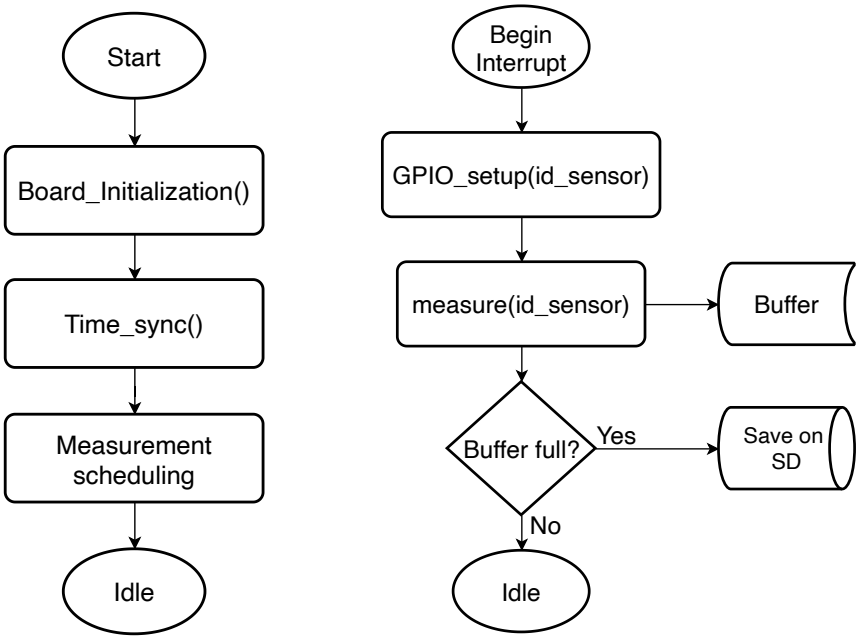
In terms of energy consumption, the Fipy board has a current consumption of less than 1 *mA*, which is relatively low compared to the CPU and PM sensor consumption. However, it may still be considered negligible in the overall energy usage of the system. To address the data storage requirements, the Fipy board is equipped with 8 MB of flash storage, which is insufficient to store PM values for long measurement periods. To overcome this limitation, a micro-SD module was added through a Pycom expansion board, and it is accessed using the SPI protocol. All the data collected during the experiment was stored in the micro-SD card in CSV format, while the internal memory of the Fipy board was solely used to store the device firmware.

Software Architecture

The Fipy development board incorporates a MicroPython [120] port, an open-source implementation of *Python 3.4*, which allows direct control over the microcontroller unit (MCU) and its peripherals using *Python 3.x* syntax. This MicroPython port includes hardware libraries specifically designed to handle the various hardware modules integrated into the development board. Additionally, it provides a communication API that facilitates IoT functionalities such as network control and remote management.

The software implementation flow is illustrated in Figure 3.14a. Initially, the `board_initialization()` process is executed, which activates the necessary WiFi and RTC peripherals for data logging. In the subsequent `time_sync()` sub-process, if the WiFi connection is established, the internal ESP32 RTC time is synchronized through the NTP protocol and then updated in the external RTC. However, if the connection or synchronization fails, the time value stored in the external RTC is utilized instead.

The last step is the `Measurement_scheduling` process, which checks if the micro SD storage is available and mounts it. Then, it sets up all sensors to their initial state, with PM sensors configured in a low-power state (fan off), and schedules measurement interruptions. It's worth mentioning that the Pycom implementation of MicroPython queues interrupts based on the order they arrive, so the interrupt routines need to be brief to ensure that the sensing periods are met.



(a) (b)

Fig. 3.14 Interrupt creation (a) and interrupt handling (b).

To handle different measurement periods for each sensor, each sensor has its own interrupt routine, as illustrated in Figure 3.14b. Due to the ESP32 having only one available UART interface, it is assigned to the specified sensor through the `GPIO_setup` process based on the `sensor_id` parameter. However, the DTH22 sensor uses a different communication interface and thus does not require this sub-process. Following this, the measurement process takes place and is stored in a memory buffer to reduce the frequency of write cycles to the micro SD card and prolong its lifespan. If the buffer reaches its maximum capacity, the measurements are then stored on the micro SD.

Methodology

The experiment comprises two phases. The first phase involves determining the coherence among the sensors and establishing a set point for assessing the impact of duty-cycle operation on the sensors' accuracy. During this phase, each sensor operates by taking measurements every second with the fan always on, and their values are compared to determine the difference and coherence between measurements. In the second phase, all four sensors are deployed in the same environment, with one

sensor designated as the reference, continuously taking measurements at a rate of one per second. The other three sensors are deployed with different measurement periods, sampling every 10, 30, and 60 seconds, based on the most significant estimation periods reported in the section 3.1. Both phases involve measuring PM_{10} and $PM_{2.5}$ values with all the sensors.

Sensor	**Phase 1**				**Phase 2**							
	Experiment 1				Experiment 2				Experiment 3			
	t_{st}	t_m	t_{off}	T	t_{st}	t_m	t_{off}	T	t_{st}	t_m	t_{off}	T
#1 (ID: 33, 34)	-	1	0	1	-	1	0	1	-	1	0	1
# 2 (ID: 35, 36)	-	1	0	1	6	3	1	10	-	3	0	10
# 3 (ID: 37,38)	-	1	0	1	6	3	21	30	10	3	27	30
#4 (ID: 39, 40)	-	1	0	1	6	3	51	60	10	3	47	60

Table 3.3 Duty-Cycle times adopted.

To ensure accurate measurements during the duty cycle, the sensor undergoes different stages over an estimation period (T). First, the sensor enters an active mode where the fan is turned on, but no measurements are taken until the stabilization time (t_{st}) is reached. Once the stabilization time is reached, the sensor performs measurements during the measurement period (t_m) and provides an average of these measurements. After that, the sensor enters an inactive state (t_{off}) where no measurements are taken, and the fan is turned off. The power consumption is lowest during the inactive state when the fan is off. The different phases of the PM measurement are illustrated in Fig. 3.15. Although the PM sensor manufacturer specified a stabilization time of 6 seconds, initial tests revealed some erroneous measurements. To avoid stabilization transients, a longer settling time of 10 seconds was used.

Measurement Location

The measurement campaign took place in Turin, Italy, in the winter of 2021. The objective of the experiment was to expose the measurement device to the common pollutants present in an urban environment. To achieve this, the device was positioned

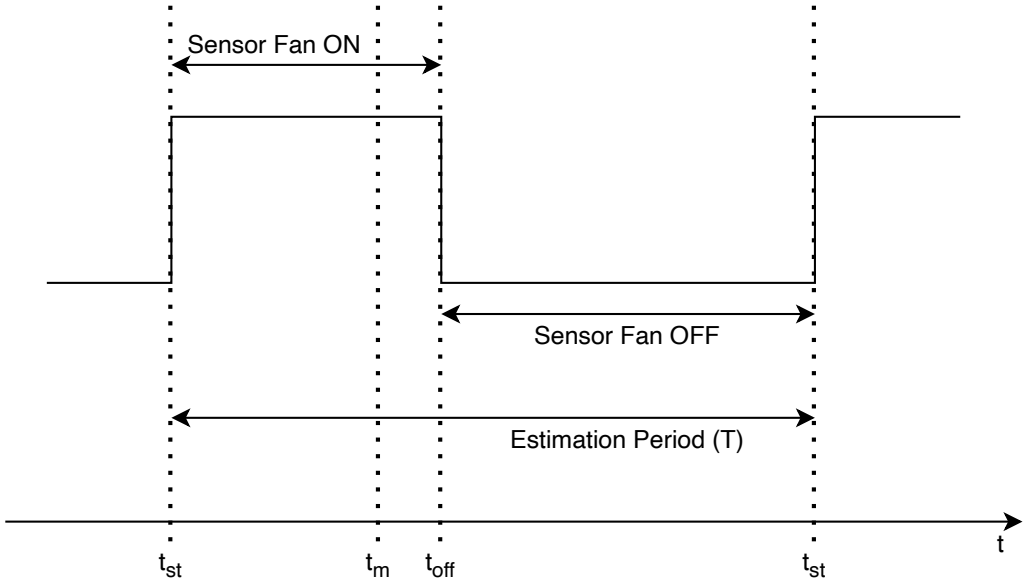


Fig. 3.15 Signal period estimation example.

at an intersection with heavy traffic flow, where various types of vehicles, including small cars, public transport, and cargo vehicles, were present (see Figure. 3.16. Due to the ongoing restrictions related to the SARS-CoV-2 pandemic, a location near a large food market was selected as it was less likely to be affected by mobility limitations imposed by local authorities.

3.2.3 Experimental Results

In this section, the data gathered in three experiments over two phases will be analyzed. The primary objective of the first phase is to establish a baseline for comparison in the subsequent phase of the experiments.

Phase 1, Experiment 1

In this phase, it was ensured that the sensors exhibited similar behavior among each other. The data obtained from the four $PM_{2.5}$(a) and PM_{10}(b) sensors are presented in Figure 3.17 after applying a filter to eliminate Gaussian noise and smoothen the signal. During the initial part of the experiment, all four sensors operated at the maximum sampling frequency of 1 sample per second to assess their mutual behavior and correlation.

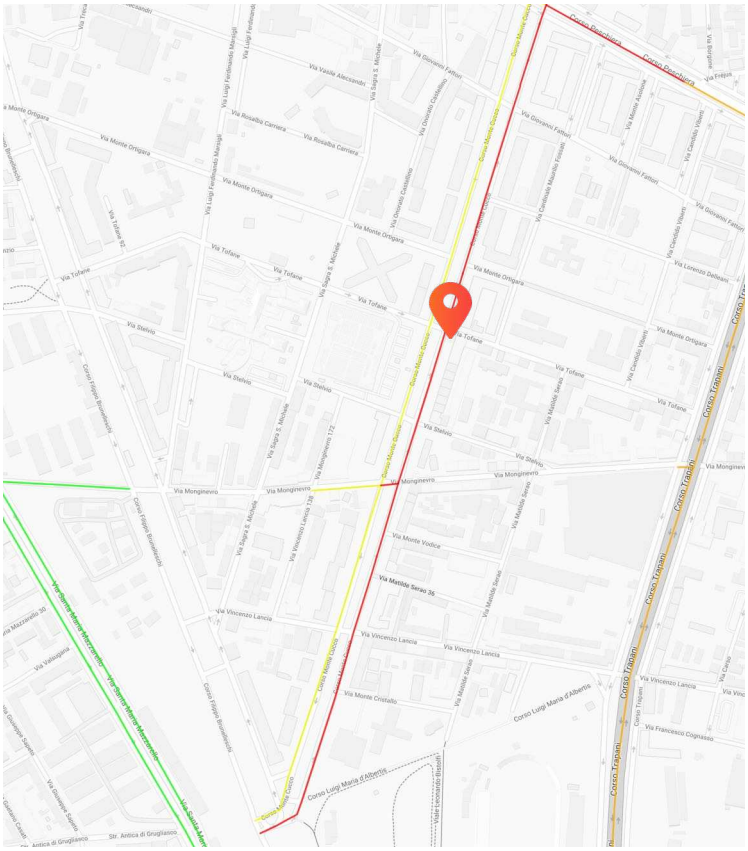
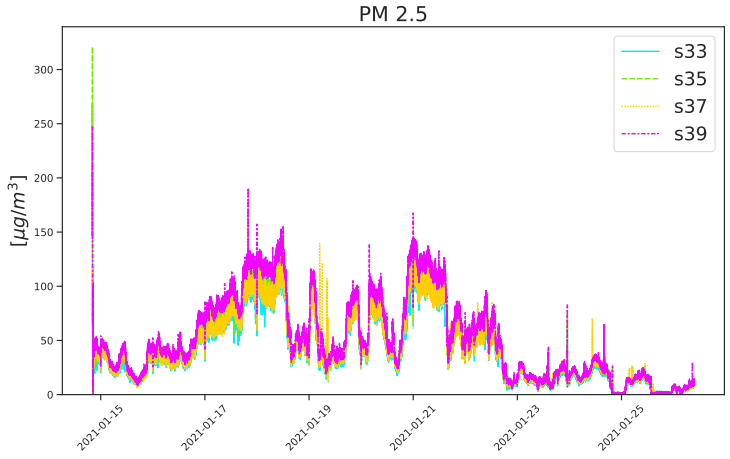
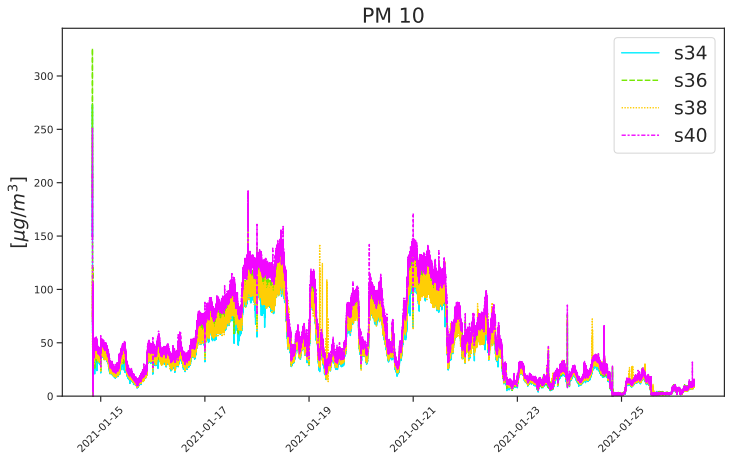


Fig. 3.16 Experiment location in Turin, Italy.



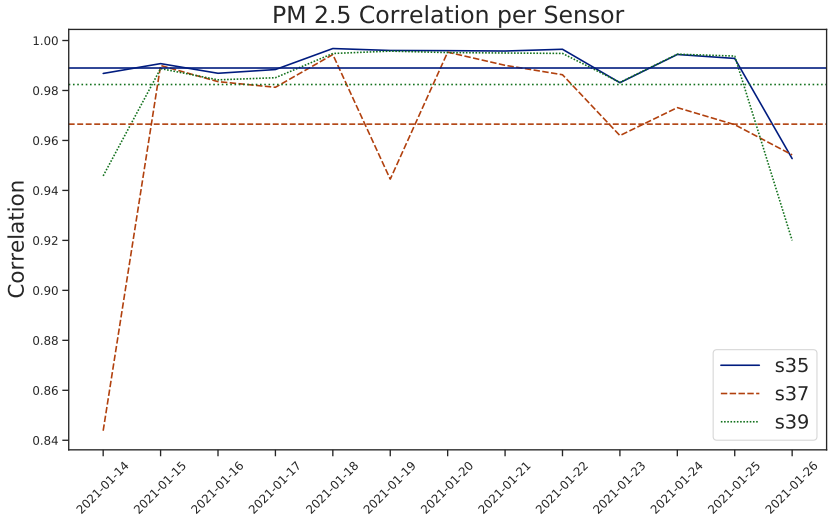
(a)



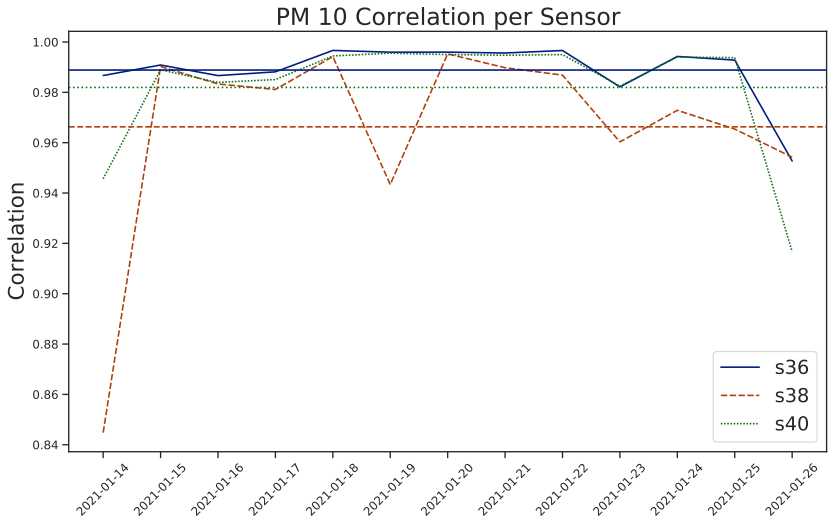
(b)

Fig. 3.17 Experiment 1: $PM_{2.5}$ (a) and PM_{10} (b) Values at one sample per second and fan always-on.

Figure 3.18 displays the correlations (including the average) between the reference sensor and the others. The average correlation factor remained consistently above 96%, indicating a high level of interdependency and reliability among these sensors.



(a)

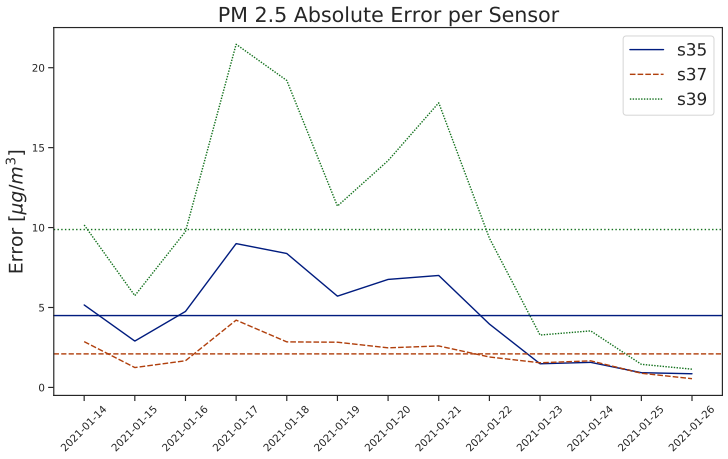


(b)

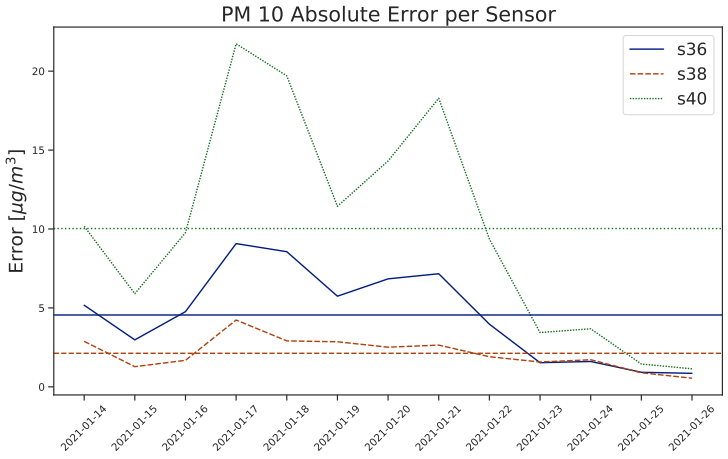
Fig. 3.18 Correlation between sensor for $PM_{2.5}$(a) and PM_{10}(b) measurement at one sample per second and fan always-on

During the corresponding time interval, Fig. 3.19 illustrates the Mean Absolute Error (MAE) of the sensors in relation to the reference sensor (including the average value). The average error does not exceed $10\mu g/m^3$, which aligns with the values

specified by the manufacturer in the datasheet and is also reasonable considering the affordable nature of these devices.



(a)



(b)

Fig. 3.19 Absolute error between sensors for $PM_{2.5}$(a) and PM_{10}(b) measurement with one sample per second and fan always-on

The provided graphs demonstrate a strong correlation among the sensors, indicating that they consistently track the PM pattern throughout the entire experiment. As the behavior remains consistent for both pollutants, we have chosen to present only the plots for $PM_{2.5}$ in the subsequent analysis.

Phase 2, Experiment 2

During this phase, the duty cycle of the sensors was modified based on the various configurations presented in Table 3.3.

In the second experiment, the sensors were configured to operate with different duty cycles. Upon powering on the sensors, a stabilization time of $t_{st} = 6s$ was allowed to pass between turning on the fan and starting the measurements. Three consecutive measurements were then collected from the sensors and averaged. The choice of a 6-second stabilization time was based on the specification provided in the sensor's data sheet, which indicated that the output becomes stable after this duration. The correlations and MAE for this experiment are presented in the correlation plot (Fig. 3.20) and the MAE plot (Fig. 3.21), respectively. The graphs indicate that the correlations and MAE values are significantly worse compared to those observed in Experiment 1. This suggests that a settling time of 6 seconds may not be sufficient to obtain reliable output from these sensors.

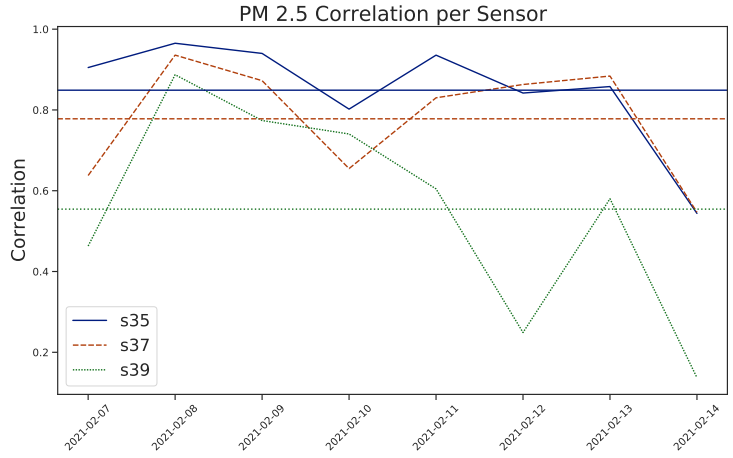


Fig. 3.20 Correlation between sensors for $PM_{2.5}$ measurement ($t_{st} = 6s$)

Phase 2, Experiment 3

In the third experiment, the duty cycle of the sensors was once again modified according to the configurations presented in Table 3.3. For this experiment, the startup time of the sensors was increased to $t_{st} = 10s$, which is 4 seconds longer than the time specified in the data sheet. The correlation plot (Figure 3.22) and the MAE plot (Figure 3.23), including daily MAE values and the average, illustrate the results

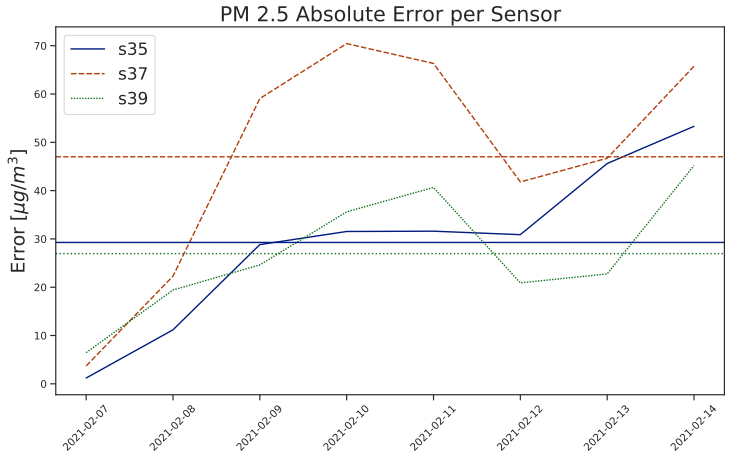


Fig. 3.21 Absolute error between sensor for $PM_{2.5}$ measurement ($t_{st} = 6s$)

obtained. The utilization of this duty cycle brings immediate improvement. The correlations between the sensors return to high percentages, and the MAE, although not optimal, is reasonable considering the reduced number of samples collected. It is worth noting that in this final experiment, the average minimum MAE is around $10\mu g/m^3$, while in the previous experiment, it was approximately $27\mu g/m^3$.

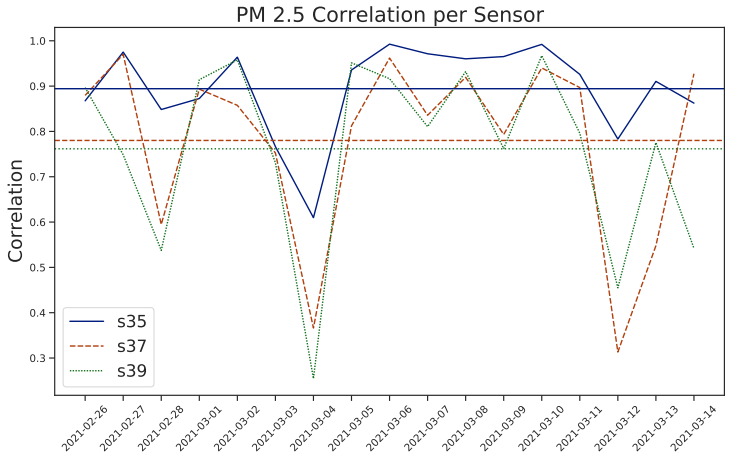


Fig. 3.22 Correlation between sensor for $PM_{2.5}$ measurement ($t_{st} = 10s$)

Energy Consumption

In order to assess the energy reduction achieved by the device introduced in this study compared to the device presented in [13], power consumption measurements

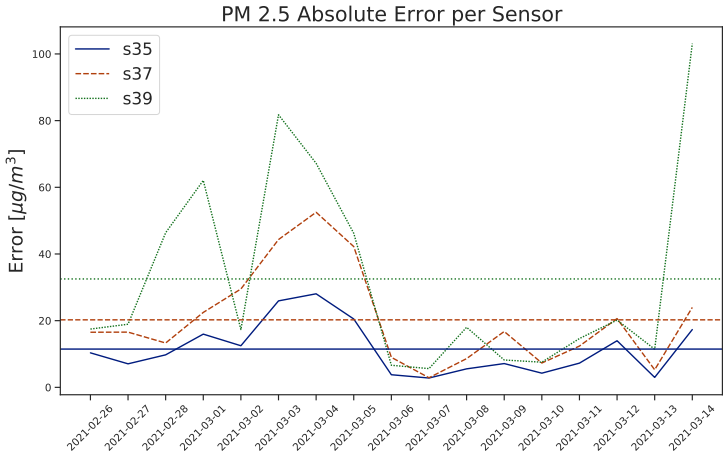


Fig. 3.23 Absolute error between sensor for $PM_{2.5}$ measurement ($t_{st} = 10s$)

were conducted using various duty cycles as outlined in Table 3.3. All sensors were configured with identical estimation parameters to facilitate a direct comparison. The findings are summarized in Table 3.4.

	Watt-Hour	**milliAmpere-hour (mAh)**
Reference Device [13] *1 sample every second*	$\approx 2,2Wh$	$\approx 460mAh$
Without Duty-Cycle *1 sample every second*	$\approx 1,8Wh$	$\approx 380mAh$
Without Duty-Cycle *1 sample every 10 seconds*	$\approx 1,8Wh$	$\approx 380mAh$
With Duty-Cycle $T = 30s$ $t_{st} = 10s$ $t_m = 3s$	$\approx 1,26Wh$	$\approx 256mAh$
With Duty-Cycle $T = 60s$ $t_{st} = 10s$ $t_m = 3s$	$\approx 1,06Wh$	$\approx 214mAh$

Table 3.4 Energy consumption benchmark (Hour average)

3.2.4 Discussion

Based on the aforementioned results, a notable observation is that employing a microcontroller significantly reduces the energy consumption by approximately 18% for performing the same tasks. Additionally, it is evident that reducing the measurement frequency to one sample every 10 seconds, while not resulting in energy savings, does yield a low MAE that falls within the manufacturer's specified tolerance range. This reduction in the estimation rate not only decreases the data volume generated but also reduces the energy consumption required for data transmission.

When the duty cycle is set to one sample every 30 seconds, it is observed that there are specific periods where the MAE and correlation between sensors are low. This behavior is evident in figures 3.24 and 3.25, which display the hourly distribution of PM concentrations during a day with a high correlation between sensors. During these specific periods, the variability between measurements decreases as the duty cycle is adjusted. Conversely, in hours with high variability, the error between measurements significantly increases across different duty cycles. Therefore, the selection of appropriate duty cycles should be based on the level of dispersion observed. Upon reviewing the data, it is found that stable concentration values are present during these periods. Consequently, depending on the dynamics of the site, it may be possible to further reduce energy consumption. Lastly, when the duty cycle is extended to T=60s, the accuracy of the measurements is also influenced by the variability of pollution concentration. On days with stable PM concentration, such as weekends with lower human activity, the correlation between sensors remains high. This mode could be employed as a low-power option in environments or during hours with minimal human traffic.

3.3 Conclusions

3.3.1 Frequency Analysis

The approach used to determine an optimal PM estimation period has revealed several key findings about the behavior of PM concentration. Firstly, it was observed that PM concentrations in urban areas exhibit varying dynamics throughout the measurement period. The analysis of stationary data indicated that PM concentration

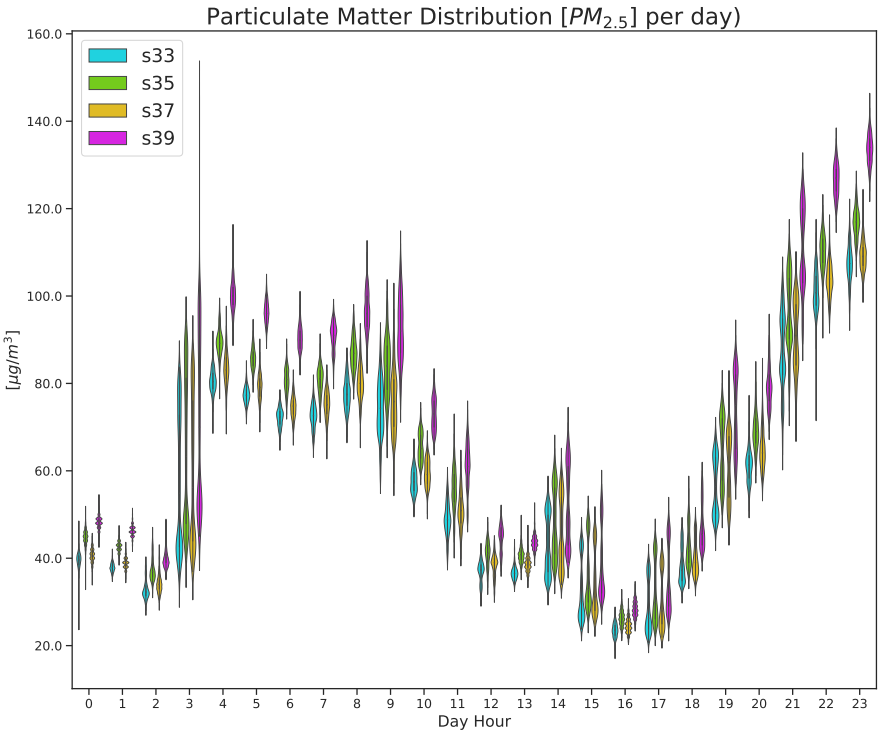


Fig. 3.24 Data distribution per hour (one day example) $PM_{2.5}$ measurement ($t_{st} = 1s$)

signals primarily consist of lower frequency components. Therefore, longer PM estimation periods are preferred to minimize the amount of data transmitted and ingested on the server-side. Additionally, it was observed that PM concentrations show reduced variability during the late hours of the night and early morning. As a result, shorter estimation periods can be employed during these periods to optimize data acquisition.

Furthermore, the steady-state nature of PM concentrations allowed for the identification of the resolution of the LSP sensors' converters. This finding emphasizes the importance of selecting conservative bandwidth criteria, particularly when the signal contains very low-frequency components, to avoid potential issues related to signal energy estimation at the 99% level.

Secondly, opting for longer PM estimation periods also enables the adjustment of the sensor's duty cycle, resulting in energy conservation, which is particularly advantageous for battery-powered IoT deployments. However, it is crucial to thoroughly examine the modification of the duty cycle to assess potential impacts on precision, such as the accumulation of dust inside the sensor and the sensor's response times.

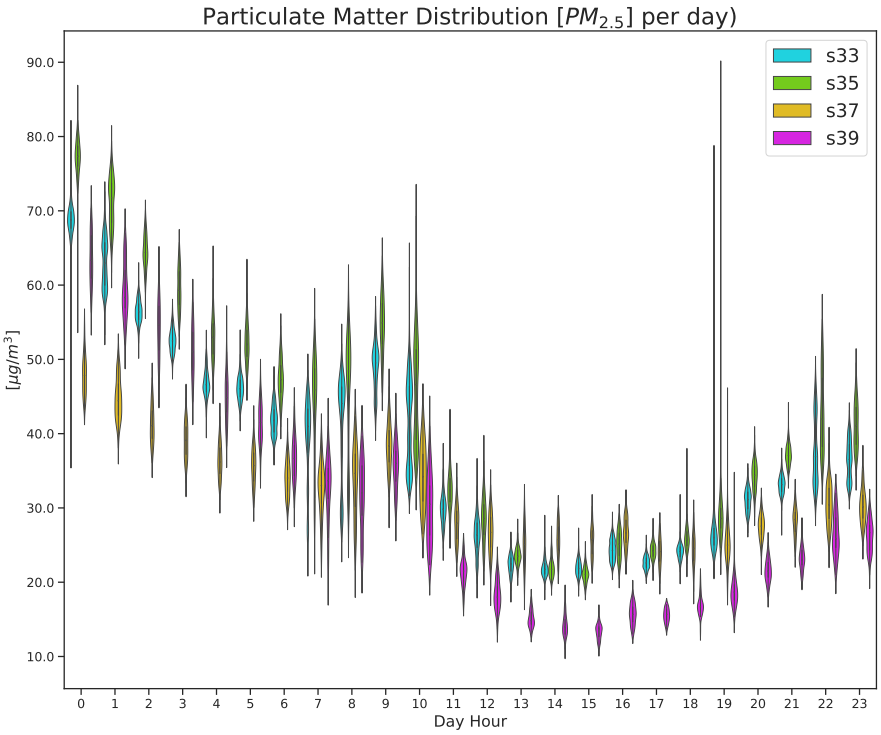


Fig. 3.25 Data distribution per hour (one day example) $PM_{2.5}$ measurement ($t_{st} = 10s$)

Thirdly, the analysis of dynamic experiments revealed an important trade-off between bandwidth precision and data reduction. The study conducted in [13] highlighted that opting for shorter estimation periods to minimize data distortion comes at the expense of scalability limitations and increased computational requirements for processing the datasets. This trade-off implies that shorter estimation times contribute minimally to PM analysis in urban environments, even when considering the response time specified by the sensor manufacturer, which is less than 6 seconds for this sensor type. Consequently, for a general estimation, periods around 30 seconds prove sufficient to capture events that lead to changes in PM levels, such as New Year's fireworks and weather phenomena. On the other hand, when analyzing PM levels in close proximity to pollution sources, a period of approximately 10 seconds provides ample information for assessing air quality levels.

In conclusion, determining a fixed estimation period for PM analysis is a challenging task given the complexity of the PM signal. The results presented in this study provide a baseline for selecting an estimation period, but they also highlight the need for further research to identify dynamic estimation rates that align with the

varying PM variability observed in urban environments. This opens up an interesting avenue for future investigations aimed at refining the estimation process based on the specific PM events of interest.

3.3.2 Data Gathering optimization

In recent years, there has been a growing interest in air pollution and its related topics, driven by increased awareness and the influence of public figures. This has led to a rise in air pollution monitoring in academia, facilitated by the reduced cost of portable Particulate Matter sensors. However, there are several engineering challenges that need to be addressed in this field, including power consumption reduction, efficient data gathering and transmission, and the aging of low-cost devices. In this section, these issues are addressed by analyzing the impact of changing the duty cycle for Light-Scattering-Particle devices. The measurements conducted in this study support the claim that the duty cycle of Particulate Matter sensors can be altered in such a way that significantly reduces the number of required samples, without compromising the amount of information obtained compared to sensors that sample continuously. This reduction in sampling frequency naturally leads to a decrease in power consumption by the sensors. Furthermore, as the sensors are not in constant operation, the moving parts inside them tend to age at a slower rate, resulting in an extended lifetime and reduced maintenance requirements.

Chapter 4

Data Transmission

In the vast landscape of the Internet of Things (IoT), numerous deployment options are available, with the specifics of the scenario dictating the best selection. Each IoT deployment embodies unique variables that necessitate careful consideration, thereby influencing the requirements that the selected technology or protocol must satisfy to maximize communication effectiveness [17, 18].

This chapter further delves into the case study of particulate matter monitoring within urban environments. The focus of the study is on addressing the challenge of data transmission by evaluating diverse data transmission schemes based on varying wireless transmission technologies, subject to the scenario. To facilitate this, we utilize the monitoring station delineated in section 3.2, which enables transmission via WiFi, Bluetooth Low Energy (BLE), LoRa, Sigfox, and LTE cellular networks.

The monitoring station design showcased in this chapter exhibits adaptability to three distinct network topologies, thus providing flexibility in connectivity. These topologies encompass Fixed Sensor Networks, wherein the stations remain stationary at a specific point; Participatory Sensor Networks, where users play an active role in data transmission while also obtaining beneficial information; and Mobile Sensor Networks, where the stations traverse multiple routes within a given area.

Further, this chapter unfolds proposals for optimizing data transmission in scenarios featuring low throughput technologies like BLE and LoRa. It also sheds light on various architectures deployed for the collection and storage of monitored values supplied by the system. These considerations highlight the complex interplay between system design, deployment scenario, and data management strategies.

4.1 Background

The scientific literature offers a wealth of information on the implementation of IoT solutions in urban settings. In particular, the measurement of air quality has attracted significant attention, leading to the development of various fixed, mobile, and personal solutions. These solutions leverage a range of wireless technologies to enable effective data collection and analysis.

4.1.1 Wireless Wide Area Networks for IoT

Cellular Networks and NB-IoT

Referring to wide-area networks (WANs), several technologies are available, with the most well-known being cellular networks defined by the 3GPP, which have been the subject of numerous previous studies. According to the investigation undertaken by [121], the researchers introduce a platform designed for measuring multiple pollutants, including PM, CO, and ozone. The platform utilizes a combination of WiFi and GPRS/UMTS (2nd and 3rd generation cellular networks) for efficient data transmission. By installing the platform on tram roofs and strategically locating it at fixed points of interest, the researchers aim to achieve extensive coverage across the city. The study focuses on addressing the spatiotemporal limitations of sensor networks by leveraging the existing infrastructure of public transportation. However, the paper lacks specific information regarding the data types, quantities, and the system's performance in transferring such data.

Similarly, other monitoring platforms rely on cellular networks, whether they are commercial networks like the one evaluated in [64], or designs that leverage the flexibility of these networks to create highly portable devices [122], enhancing spatial resolutions and controlling position through GPS systems [123, 124]. Some of these studies propose measurements in fixed locations[72, 125, 126], such as installations on vehicles traversing the city [127–129], or through devices carried by users [130].

While cellular networks, in their various generations, have improved their performance [131], they are not always the most efficient option for IoT systems. Particularly concerning aspects of power consumption, latency, and quality of ser-

vice. The latest generations (LTE and later) offer IoT support within their architecture (NB-IoT). These networks address the weaknesses of previous generations, as their usage in IoT is subject to third-party infrastructure, which improves coverage options in various cities based on service provider availability. This technology even demonstrates superior performance for mobile devices compared to other technologies, making it ideal for urban environments when seeking mobile deployments with extensive coverage, high-volume data transmission requirements, low latency, and quality of service (QoS) policies [131, 132]. In contrast, the operational costs associated with NB-IoT are higher, as they require increasing operational expenses with the number of devices and transmitters, resulting in lower cost-efficiency relative to other wireless wide-area alternatives [133]. Additionally, NB-IoT still consumes more power than other wide-area wireless alternatives.

Low-power wide-area networks (LPWANs) offer an alternative to cellular networks and have emerged as a prominent solution for long-range communications in modern IoT systems. LPWANs exhibit remarkable capabilities in facilitating communication between sensors that are geographically dispersed, even over extensive distances of up to hundreds of kilometers. This is achieved while operating on energy-efficient battery technologies, which enables the deployment of IoT applications and services previously unfeasible. The advent of LPWANs has paved the way for a wide range of innovative applications and services in the IoT landscape.

The LPWAN technologies possesses distinctive attributes that distinguish it from other wireless technologies as was presented in [132, 131]. Firstly, LPWAN offers an impressive operating range, spanning from a few kilometers in urban areas to over 30 km in rural environments. This enables efficient data communication in challenging locations, including indoor and underground settings. Secondly, LPWAN is designed with low power consumption in mind, allowing transceivers to operate on small, cost-effective batteries for extended periods, often up to 10 years. Thirdly, LPWAN solutions are characterized by their affordability. The implementation of simplified, lightweight protocols reduces complexity in hardware design, resulting in lower device costs. Moreover, the long-range capabilities of LPWAN, combined with a star topology, minimize the need for costly infrastructure. Additionally, the use of license-free or licensed frequency bands contributes to reduced network expenses. Lastly, LPWAN is optimized for low data rates, enabling cost-effective communication for small amounts of data, often costing only a few Euro cents per

month. LoRa/LoRaWAN and Sigfox are the most commonly used technologies in the IoT environment, a brief comparison of their specifications is showed in table 4.1

	NB-IoT	**Sigfox**	**LoRaWAN**
Range(Km)	1 (urban) to 10 (rural)	10 (urban) to 40 (rural)	5 (urban) to 20 (rural)
Channel BW(Hz)	200k	100	250k and 125k
Max. Data-Rate(bps)	200K	100	50K
Frequency(MHz)	LTE-Bands	433, 868, 915	433, 868, 915
Bidirectional	Yes Half-duplex	Limited Half-duplex	Yes Half-duplex
Max.payload(bytes)	1600	12 (Upload) 8 (Download)	243
Maximum msg/day	Unlimited	140 (Upload) 4 (Download)	Depends on server and data rate

Table 4.1 LPWAN Comparison

Sigfox

Sigfox is an LPWAN network operator that provides a comprehensive IoT connectivity solution based on its proprietary technologies [134]. The cognitive radio base stations deployed by Sigfox establish a connection with the backend servers through an IP-based network. Devices connected to these base stations utilize Binary Phase Shift Keying (BPSK) modulation on a sub-GHz ISM band carrier with an extremely narrow bandwidth (see table 4.1). This technology optimizes frequency spectrum utilization by employing an ultra-narrow band, resulting in minimal noise levels, reduced power consumption, high receiver sensitivity, and cost-effective antenna design. However, it achieves a maximum throughput of only 100 bits per second (bps), and as a proprietary technology, Sigfox requires the payment of fees for communication services[135].

Although Sigfox is highly energy-efficient, its use in air quality monitoring is limited due to its low transmission rates. Sigfox is more suitable for sending averaged values over time or variables that exhibit slow changes. In the literature, we observe a few implementations primarily focused on monitoring variables such as temperature, humidity [136], and, in some cases, atmospheric pressure [137].

Initially supporting only uplink communications, Sigfox has evolved into a bidirectional technology with significant link asymmetry. Downlink communication, allowing data transfer from base stations to terminals, is constrained to occur after uplink communication and is limited to a maximum of four instances per day. This limitation restricts the base station's ability to acknowledge every uplink message [138].

LoRA/LoRaWAN

LoRa and LoRaWAN are wireless communication technologies that have been developed specifically for IoT applications, providing long-range and low-power connectivity for IoT devices. LoRa, developed by Semtech Corporation [139], is a physical layer modulation technique that utilizes chirp spread spectrum modulation. This modulation allows for long-range communication by taking advantage of low data rates, high receiver sensitivity, and resistance to interference. LoRa operates in the unlicensed radio spectrum (see Table 4.1) and can achieve a link budget of 157 dB.

LoRaWAN, on the other hand, is a protocol built on top of LoRa technology, providing the network layer for communication between LoRa devices and network servers [139]. It facilitates secure and bidirectional communication, allowing IoT devices to transmit data to a centralized network server and receive commands or acknowledgments in response. LoRaWAN also incorporates features like adaptive data rate [140], enabling devices to dynamically adjust their transmission rates based on the quality of the radio link. This helps optimize communication efficiency while ensuring reliable connectivity for IoT devices.

LoRa and LoRaWAN provide a significant advantage in the field of IoT due to their long-range connectivity capabilities and low power consumption. With LoRaWAN, large areas can be covered efficiently through the use of a single gateway, enabling cost-effective and scalable deployments. In a typical LoRaWAN network, a star-of-stars topology is employed, where multiple end-devices establish communication with gateways. The gateways then transmit the collected data to a centralized network server. Subsequently, the network server processes the data and forwards it to the application server or cloud for further analysis and action.

The potential of LoRa/LoRaWAN for air pollution monitoring has been extensively explored in the literature. A summary of some proposals can be found in [141]. In this work, various approaches are presented, such as the one introduced in [142], which utilizes high-resolution and expensive sensors. These precise sensors generate data at low rates, making them suitable for transmissions under this technology. However, due to the cost of the sensors, this proposal is hardly scalable in an urban environment.

Other researchers have developed prototypes using low-cost sensors. These prototypes evaluate different types of air quality sensors and conduct proof-of-concept tests on transmission using various MCUs [143–145], while assessing both power consumption and latency. However, these prototypes do not specify transmission rates, payload size, message rates, and network performance. Some prototypes are evaluated in fixed positions for monitoring a particular geographic sector [146]. Additionally, others focus on analyzing the obtained data without an in-depth analysis of LoRa's performance [147–149].

Furthermore, other studies have deployed multiple sensors to create a distributed system using fixed stations [150], as well as mobile solutions. Among the mobile solutions, there are limitations on mobility traceability, restricted to low speeds and slow trajectories due to the low message rate provided by LoRa [151]. Deployments for vehicles have also been explored, enabling mobile monitoring systems using automobiles [152], or autonomous unmanned aerial vehicles (UAV), where the concept of mobile transmission in these vehicles is analyzed [153, 154]. However, only [152] presents an analysis of LoRa's performance by examining the packet delivery ratio (PDR).

Finally, in [155], a feasibility analysis is presented for air quality monitoring, considering transmission performance based on distance, PDR, and signal strength within an urban environment. Additionally, the potentiality is analyzed according to the amount of data to be used, although not in depth.

4.1.2 Local Area Networks for IoT

Wireless Fidelity (WiFi)

WiFi technology encompasses the various IEEE 802.11 standards [156], which define the different generations and specifications of this wireless technology [157]. It has become an integral part of individuals' lives, being incorporated into personal, household, and even enterprise wireless devices. Within the IoT context, WiFi assumes a crucial role in enabling communication among interconnected devices. Reliable and secure wireless connectivity is essential for IoT devices to exchange data, transmit commands, and collaborate within a network. WiFi infrastructure serves as a necessary foundation for supporting IoT deployments, allowing devices to establish communication with one another and connect to the internet.

The widespread adoption of WiFi has enabled various IoT solutions to be supported by it. It is common for embedded development platforms to include WiFi connectivity [158]. WiFi's potential for IoT is extensive. It facilitates the integration of diverse devices into IoT networks, encompassing smart home devices, wearables, industrial sensors, healthcare devices, and more. The ubiquity of WiFi networks simplifies the adoption and integration of IoT devices, fostering an ecosystem where devices can communicate, share data, and collectively contribute to intelligent decision-making [159].

Furthermore, the strength and high data transfer speeds of WiFi make it well-suited for IoT applications that involve substantial amounts of data. IoT devices often generate a large volume of data that requires real-time processing, analysis, and action. WiFi's ability to handle extensive data streams facilitates timely data transmission, supporting applications such as real-time monitoring, video surveillance, and predictive maintenance in industrial environments [160].

As WiFi technology continues to evolve, with the emergence of standards like WiFi 6 (802.11ax) and WiFi 6E, the capabilities of WiFi for IoT are further expanded. These standards offer increased capacity, faster data transfer rates, reduced latency, and improved efficiency, enhancing the overall performance of WiFi networks and enabling smoother integration of IoT devices. [161]

Although WiFi is not among the lowest power-consuming technologies, several studies demonstrate its favorable results compared to other low-power technolo-

gies, as shown in [162, 163]. Additionally, other research showcases how energy consumption can be optimized in IoT devices [164].

Several surveys in the literature on IoT, specifically in the context of air quality monitoring [69, 141], present various solutions that incorporate WiFi for communications, either as the primary transmission mechanism [13, 91] or as an alternative transmission method [87, 165, 128].

As another alternative, options for low-power (802.11ah [166]), wide-coverage WiFi standards (such as WiMax [167]) have also been explored, although they have not been widely adopted. Therefore, in environments where access to a WiFi network that meets the solution's requirements and provides internet connectivity is available, WiFi proves to be an economical and practical alternative for IoT deployments. In the case of air quality monitoring, systems like the ones presented in Chapters 2 and 3 offer an efficient solution for transmitting the measurements.

In summary, WiFi emerges as a valuable solution for addressing connectivity challenges in air quality monitoring IoT solutions. Its wide adoption, mobility features, high-speed data transfer capabilities, and robust security protocols make it an ideal choice for connecting air quality sensors and facilitating real-time data exchange when a WiFi coverage is available.

4.1.3 Personal Area Network

802.15.4

WSNs based in the standard IEEE 802.15.4 [168] is another wireless technology that have attracted considerable interest in both research and industry due to their wide range of potential applications. These networks utilize small, self-contained devices known as wireless sensors to collect data from the environment and transmit it wirelessly to a central location for further analysis. The versatility of 802.15.4 enables their deployment in diverse areas, including environmental monitoring, healthcare, and smart cities.

The IEEE 802.15.4 is a MAC standard extensively utilized, with popular implementations like ZigBee [169] and 6LoWPAN[170]. Among these, 802.15.5 is specifically designed to cater to the demands of low-power and low-data-rate wireless communication in WSN. This standard offers a flexible and reliable communication

protocol that optimizes resource utilization and minimizes power consumption. It operates in unlicensed frequency bands, including 2.4 GHz and 868/915 MHz, making it well-suited for various applications and deployment scenarios [171].

WSN nodes are commonly powered by batteries and often require extended periods of operation without the need for battery replacement or recharging. To address this challenge, the standard incorporates power management mechanisms, such as duty cycling and sleep modes. These mechanisms enable nodes to conserve energy by periodically entering low-power states when they are not actively involved in transmitting or receiving data. By utilizing low transmission power and operating at low data rates, the system achieves a range of a few meters (up to 100 meters), which is more efficient compared to technologies like WiFi.[135] By adopting this energy-conscious approach, the network can prolong its overall lifetime and support long-term operation in environments with limited resources [110].

IEEE 802.15.4 is also characterized by its capability to support ad-hoc networking, which is highly advantageous in wireless sensor network (WSN) deployments. These networks are often deployed in dynamic and distributed environments where the network topology can change frequently. The ability of IEEE 802.15.4 to self-organize and support a multi-hop network structure allows sensors to communicate with each other in a mesh-like manner. This enables WSNs to cover large areas without the need for extensive network infrastructure. Furthermore, this self-organizing capability provides a self-healing mechanism that ensures network connectivity even in the face of node failures or environmental obstacles [171]. This enhances the overall reliability and resilience of the WSN, making it well-suited for diverse and challenging deployment scenarios.

When it comes to security, the various implementations of the IEEE 802.15.4 standard offer mechanisms that effectively address the need for data integrity and security [172]. These mechanisms encompass features such as encryption, authentication, and access control. By incorporating these security measures, the standard addresses the challenges associated with transmitting sensitive data in environments that may be hostile or untrusted. This ensures that the data remains protected and secure throughout its transmission within the network.

The literature surrounding communication systems based on IEEE 802.15.4 is extensive and encompasses a wide range of applications [173, 174], as well as exploring the capabilities of this standard in terms of energy efficiency [110],

routing protocols [175], localization algorithms, quality of service, among others. Researchers have also investigated the performance evaluation and optimization of WSN deployments, considering factors such as network scalability, reliability, latency, and throughput [176].

This technology has also attracted the attention of researchers in the field of air quality monitoring. Several studies have utilized this technology for transmitting pollutant measurements. In [69], a review of various network deployment types is presented, categorizing them as static sensor networks, participatory sensor networks [177], and vehicular sensor networks. From [69] and other literature reviews, it is evident that the majority of deployments using this technology are observed in static sensor networks due to the complexity associated with ensuring packet delivery and preventing routing loops in a mobile system.

Examples of static monitoring can be seen in the work presented by the authors in [178], where they propose an indoor air quality monitoring system using ZigBee. Other solutions, such as the one presented in [179], propose a star topology network for gas monitoring. Another alternative is presented in [116], where the authors propose a multi-hop distributed network system for pollution measurement, incorporating a data aggregation algorithm to optimize transmission.

Other systems more oriented towards urban environments are presented in [180], where various sensors are deployed around a street and roundabout for pollutant measurement and spatial distribution analysis. This deployment takes advantage of the routing capabilities of the devices to achieve wide coverage without requiring the deployment of infrastructure. To overcome limitations in this aspect, IEEE 802.11ah is profiled as an option to achieve greater coverage [181].

In the mobile environment, several studies have been conducted using public or private vehicles [182–184]. In other investigations, the use of UAVs for air quality monitoring has been explored [86]. In these latter schemes, since 802.15.4 has a limited range, the implementation of other communication systems to ensure data transmission or the determination of synchronization points has been necessary, affecting temporal resolution.

Bluetooth

Bluetooth and Bluetooth Low Energy (BLE) [185] are wireless communication technologies that have gained significant attention in various domains. They are widely used in consumer electronics such as smartphones, computers, wearables, automobiles, industrial devices, among others. This widespread adoption enables a high potential for interoperability between computing systems and IoT devices.

Bluetooth is a cost-effective wireless communication standard designed for short-range data transmission between devices. It operates in the unlicensed 2.4 GHz frequency band, ensuring reliable and secure connectivity. With its simplicity and convenience, Bluetooth provides an efficient solution for device connection and data exchange, making it well-suited for a wide range of applications, including wireless sensor networks and IoT deployments [17].

Bluetooth Low Energy (BLE), also known as Bluetooth Smart is added in the Bluetooth 4.0 definition. It is a power-efficient variant of the Bluetooth standard. BLE is specifically designed for applications requiring low power consumption, making it ideal for battery-powered devices and IoT applications. It provides a balance between data transfer rate, range, and energy efficiency, enabling long-lasting operation in resource-constrained environments[186].

The performance characteristics of Bluetooth and BLE are influenced by various factors such as data transfer rate, range, power consumption, and connection stability. Bluetooth generally provides higher data transfer rates, whereas BLE prioritizes energy efficiency and operates at lower data rates. Both technologies encounter challenges associated with interference from other devices operating in the congested 2.4 GHz frequency band. Nevertheless, advanced frequency hopping techniques are employed to address interference issues and improve overall reliability.

Both technologies have an average range of a few meters, approximately 20 meters, without experiencing significant degradation in communication. However, the actual range can vary depending on factors such as antenna gains, receiver sensitivity, physical layer (version of the standard), interference, path losses, and transmission power [187].

For smart cities, the utilization of crowd sensing and participatory sensing technologies presents immense potential for the implementation of air quality monitoring systems [68]. By employing portable solutions carried by individuals or deployed

within communities, the spatial resolution of air quality measurements can be significantly improved[69].

These solutions often utilize mobile applications that enable user engagement with the monitoring system. The application collects, adjusts, and transmits data to an online data collection point, where it undergoes processing and visualization for user accessibility and reference. Research studies, such as those presented in [87, 188, 165], have explored various approaches for air quality monitoring. In some of these studies, the transmission is facilitated through a mobile device equipped with a customized application designed specifically for this purpose. Additionally, other systems utilize smartphones and Bluetooth connectivity as a secondary communication method, either as an alternative means of data transmission or for reviewing the collected measurements [94].

Another approach involves the implementation of participatory sensing models, where strategically placed sensors are interacted with by users through their smartphones to query, collect, and transmit information [177, 189]. The implementation of crowd sensing and participatory sensing technologies in air pollution monitoring systems shows great potential in the context of smart cities. These approaches, which involve portable devices and active user participation, offer advantages in terms of expanding spatial coverage, improving data accuracy, and ensuring timely measurements of air quality. The integration of mobile applications and user interactions not only enhances the effectiveness of environmental monitoring but also promotes citizen engagement, fostering inclusivity and empowerment.

4.2 System Architecture

To support the diverse deployments, this research proposes a common architecture for the evaluated scenarios. As depicted in Figure 4.1, the main component of this architecture is the air quality monitoring stations, which offer connectivity flexibility, allowing the same design to be applied in multiple scenarios. Data transmission is primarily established using the Message Queuing Telemetry Transport (MQTT) v3.1.1 protocol.

MQTT is an efficient messaging protocol specifically designed for IoT applications. It employs a publisher-subscriber model, enabling devices to exchange data

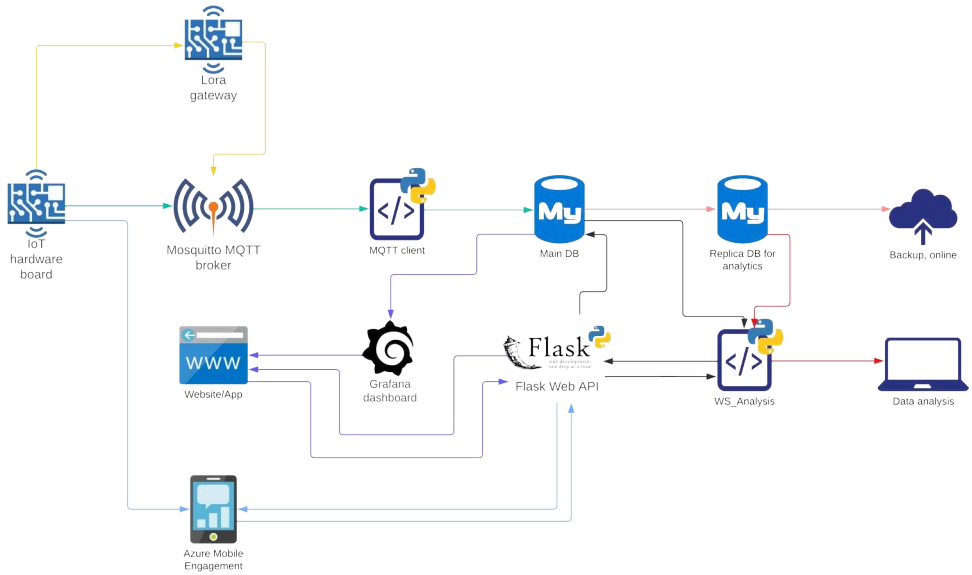


Fig. 4.1 System architecture

via a central broker. Due to its minimal data overhead and compatibility with diverse network connectivity options, MQTT is an ideal choice for resource-constrained devices and networks commonly found in IoT environments[190].

In the case of LoRaWAN network support, MQTT enables the system to utilize gateways that act as bridges between the device and an IP-based network. These gateways are controlled by a LoRaWAN server, which manages the entire network, including data encryption, protection, and handling of duplicate packets received by the gateways. When utilizing IP-based transmission such as Wi-Fi or NB-IoT, MQTT transmission occurs directly over a TCP socket.

An MQTT client that subscribes to designated MQTT topics for each sensor per monitoring stations obtains the data. It decodes binary-format data, if present, and then stores all received information in the central database.

For monitoring stations that utilize Bluetooth/BLE connection, data transmission occurs through a web application, which redirects the data to a REST server (via Flask Web API) for storage in the database (see Figure 4.2. The application not only allows data uploading but also facilitates data querying and visualization through

REST queries to the server. The server retrieves the stored data, applies calibration and filtering, and presents it in a visual format.

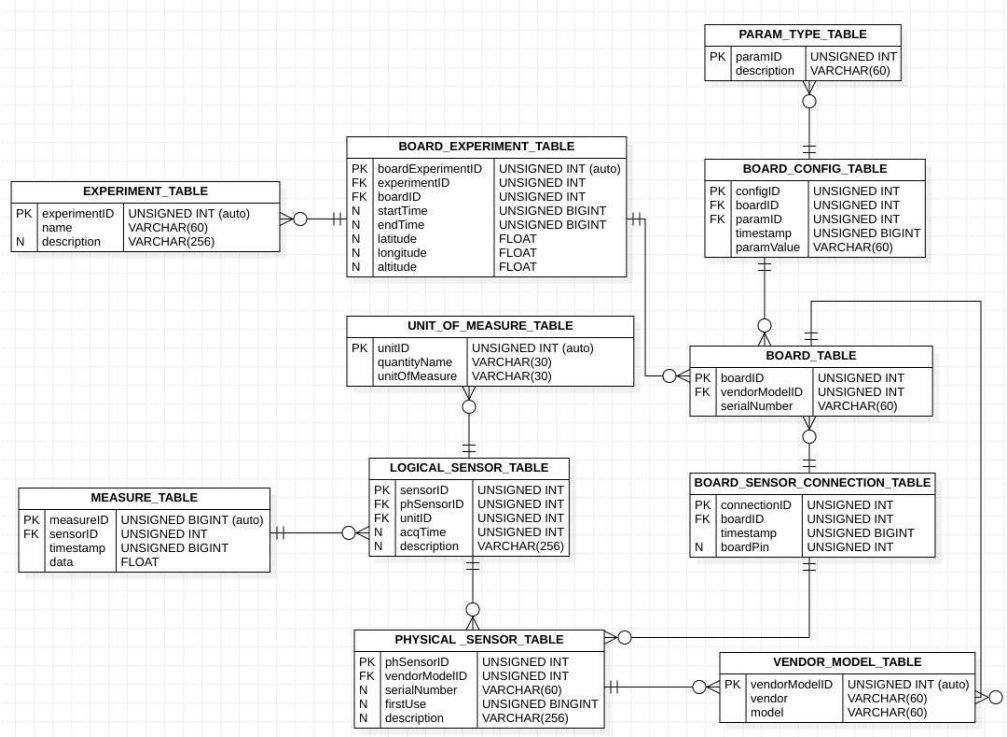


Fig. 4.2 Database relational schema

Furthermore, the architecture includes calibration tasks (WS_ANALYSIS) that apply sensor calibration to enable the visualization of post-processed data. It also allows data calls to the database for further data analysis in associated research (refer to Chapter5).

4.3 Air Monitoring Stations Overview

4.3.1 Hardware Overview

The air monitoring station used in this chapter is the same described briefly in the section 3.2.2. Internally, the air monitoring station consists of four *HPMA115S0-XXX* sensors for measuring concentrations of particulate matter (PM_{10} and $PM_{2.5}$). A *BMP280* atmospheric pressure sensor, a real-time clock for synchronizing measurements with the GMT time, a *DTH22* relative humidity and temperature sensor , and

a *P1010D* GPS unit for determining the station's geolocation (see Figure 4.3). The *DTH22* sensor and the GPS unit are installed outside the sensor enclosure to avoid interference during sensor stacking (during calibration scenarios) in the case of GPS and to prevent measurement errors caused by heat generated by other sensors and the processing unit inside the case. The Pycom implementation of MicroPython utilizes interrupt queuing based on arrival order. Therefore, it is important to keep the interrupt routines concise to ensure that the sensing periods are adhered to. Consequently, each sensor is configured with its specific sampling frequency, and when a timer reaches its expiration, a routine is triggered to retrieve the measurement value generated by the sensor.(see Figure 4.4

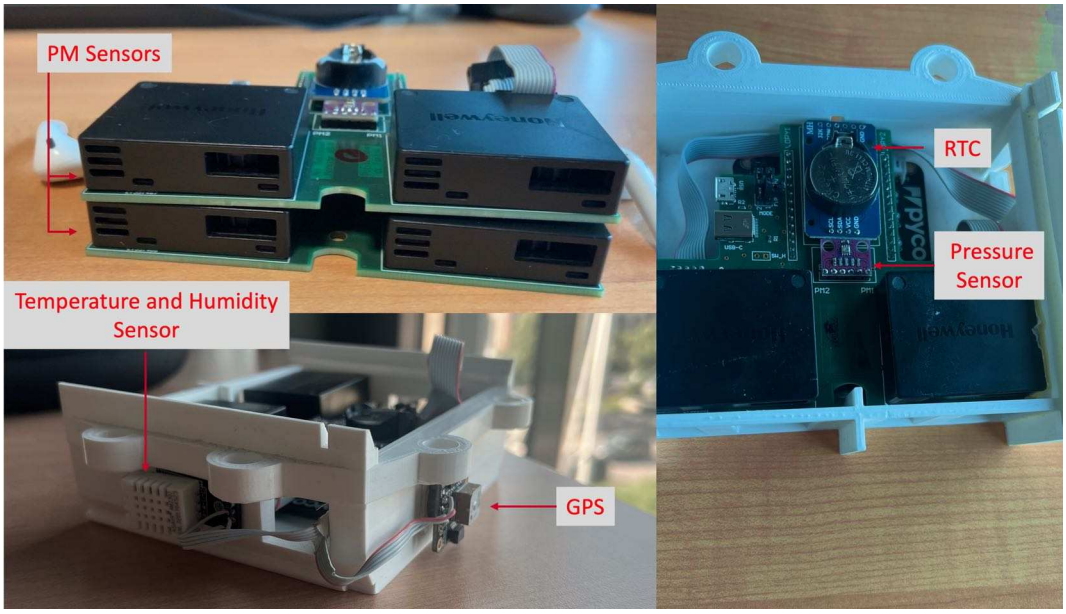


Fig. 4.3 Board design, sensors, and final measuring station

As computing units, two models of development boards based on the ESP32 microcontroller are used. The ESP32 [119] is a low-cost, low-power processor that features dual cores and an ultra-low-power co-processor. Its main feature is the integration of Wi-Fi (802.11 b/g/n at 2.4 GHz) and Bluetooth (v4.2, BLE). Along with this processor, other communication modules are embedded based on the development board version:

- **Pycom LoPy4**: This development board includes a module for SigFox and LoRa communication [191].

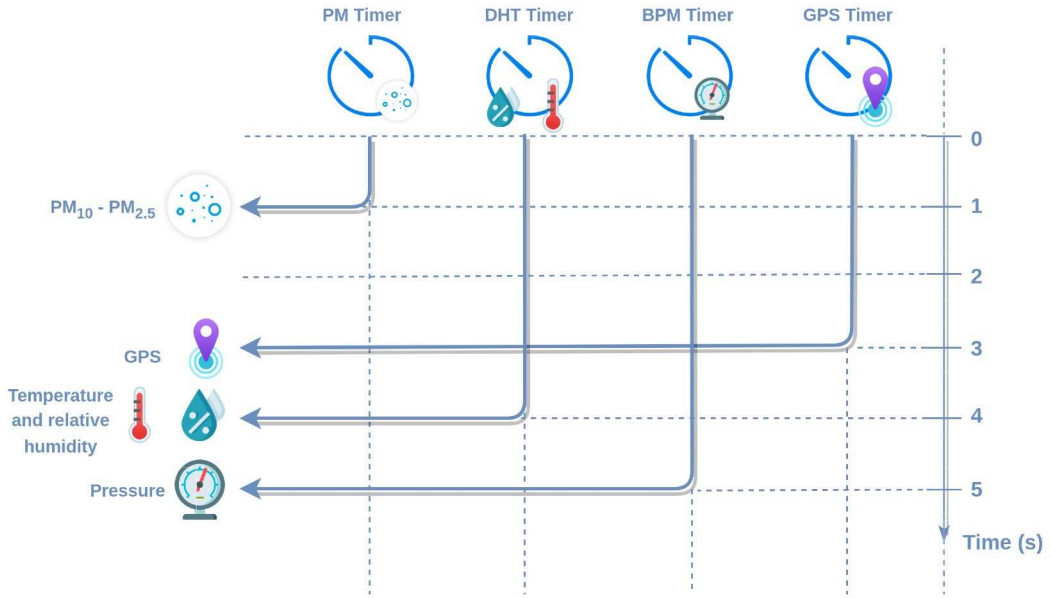


Fig. 4.4 Sensor Measurement Scheduling. When a timer expires, a routine retrieves the measurement value produced by the sensor.

- **Pycom FiPy**: This development board includes a module for SigFox and LoRa communication, an additional module for mobile connectivity through LTE CAT M1/NB-IoT networks, and supports a nano SIM [118].

Both boards are supported on an expansion board (PCB) that includes GPIO ports, programmable led and push buttons, USB port for serial communication or for module programming, and a slot for Micro-SD card to save all the measurements. For the installation of the sensors and compatibility with both development board models, a PCB and a corresponding enclosure are designed to protect the monitoring station for outdoor operation. The placement of the PM sensors, along with the PCB and enclosure design, follows the guidelines described by [192] to prevent airflow feedback that may affect precision. (see Figure 4.5

4.3.2 Software Overview

The FiPy/Lopy4 development boards works under a custom implementation of Micropython [120]made by Pycom. MicroPython is based in the *Python 3.x* syntax, and includes a small subset of the Python standard library. This implementation

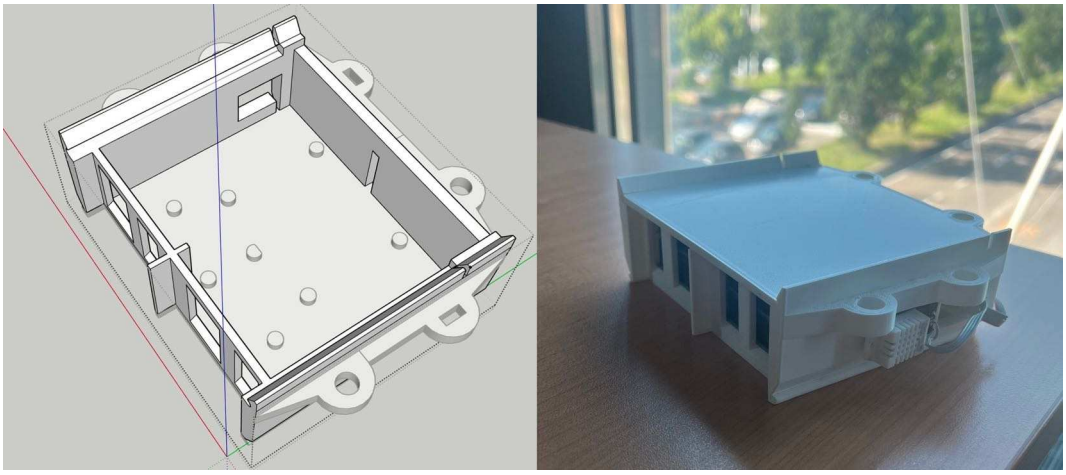


Fig. 4.5 The stackable modular 3D printed case

comes with all the necessary libraries to handle the different networks embedded on the boards.

The Pycom implementation provides an interactive command prompt (REPL) for executing commands and running scripts stored in the system memory. This REPL can be accessed via USB connection or remotely via WiFi using the Telnet protocol. The MicroPython scripts that define the firmware behavior are stored in memory and can be created, modified, or deleted using the manufacturer-supplied plugin or passive FTP.

The software behavior follows the description in chapter 3.2.2. However, to simplify device configuration, two JSON files are used to contain the device settings for specific deployments:

- CONFIGSENSORS.JSON: Contains information about the board type (LoPy or Fipy), state configuration (On/Off), sampling period, and ID for each sensor. Additionally, it defines the buffer size for data transmission/storage, transmission type, and data storage/transmission format (binary/text).
- NETCONFIG.JSON: Defines the ID of the monitoring station, security credentials for FTP/Telnet access, and transmission parameters based on the chosen technology:
  - WiFi:
    * WiFi mode (Access Point or Client)

    - Known networks (SSID and password)
    - MQTT credentials for connecting to the MQTT broker
  - LoRA:
    - LoRaWAN authentication credentials
    - ADR usage (On/Off)
    - Data rate
    - Sleep time
  - BLE:
    - Device name
    - Service UUID (unique numeric ID) for advertising the service
    - GATT (General Atribute Transfer Protocol) parameters

4.4 Deployment Area

The case study presented in this research is conducted in the metropolitan area of Turin, located in northern Italy. The air quality monitoring is carried out by ARPA using a network of fixed stations. The location of each station depends on the level of urbanization (urban, suburban, and rural) as well as proximity to potential sources of pollution (see Figure 3.11). However, these stations do not provide sufficient spatial and temporal resolution in several areas of the city. This study examines various deployments to enhance the resolution.

4.5 Deployment Scenarios

The complexity of urban environments requires different types of deployments based on the specific objectives of observation and analysis. Similar to the classification presented by [69], the following network taxonomy is proposed:

- Fixed Sensor Networks
- Participatory Sensor Networks
- Mobile Sensor Networks

4.5.1 Fixed Sensor Networks

In this case, sensors are placed at specific points where there is a particular interest in monitoring a specific environment or the behavior of particulate matter. Examples of this type of deployment are presented in Chapter 2, showcasing both indoor and outdoor environments.

In fixed sensor deployments, energy constraints are typically low, as there is usually a nearby power source or energy harvesting methods (such as solar panels) can be employed. The stationary nature of these sensors also reduces weight and size constraints, allowing for a greater number of sensors to be deployed. In terms of networking, considering a fixed position allows for the selection of a specific communication technology since the conditions remain stable, ensuring a reliable connection.

However, this deployment model also presents challenges. Determining the points of interest in a city requires careful attention to achieve effective observability of pollutants. Additionally, the spatial-temporal resolution is limited, and addressing this issue would require deploying a larger number of monitoring stations. In terms of data, if the conditions remain stable, there is a lower amount of relevant information to transmit. From Communication standpoint, these systems typically rely on mobile technologies, which increase costs as the number of monitors grows. Using alternative technologies would necessitate customizing and deploying additional networks. Finally, maintenance would require mobile personnel to carry out these tasks.

Two deployment scenarios are proposed in this case: Calibration, and urban monitoring using LoRa.

Calibration Scenario

This essential stage, prior to conducting measurements in the city, is crucial for the utilization of low-cost sensors. In this scenario, the sensors are colocated with a professional monitoring station operated by ARPA in Rubino. The sensors are deployed for various periods to create the calibration model using the methodology proposed in [13].

In this strategy, particulate matter is measured every second, which necessitates high throughput communication. Networks such as LoRaWAN or Sigfox are discarded due to their limited throughput. Therefore, the alternative for this scenario is cellular network connectivity.

While each board could transmit using this technology, deployment costs increase. Hence, for this scenario, a Wi-Fi access point is utilized to consolidate the monitoring stations into a single Wi-Fi network. It redirects MQTT messages to the architecture proposed in Section 4.2. The utilization of a gateway reduces deployment costs as a single connection per reference station enables communication with all the low-cost stations. Additionally, a Raspberry Pi is installed as a proxy, serving as an access point for each station. This setup allows FTP and REPL access via Telnet to the boards, facilitating remote management. While the Telnet and FTP protocols, though insecure, are integrated into the MicroPython manufacturer's firmware and utilized for remote management, their application in this research primarily stems from convenience and is not recommended for deployment. Currently, access through these protocols is exclusively confined to the encrypted Wi-Fi network (WPA2) established between the Proxy Server (Raspberry Pi) and the monitoring stations. A summarized architecture for this scenario is presented in Figure 4.6.

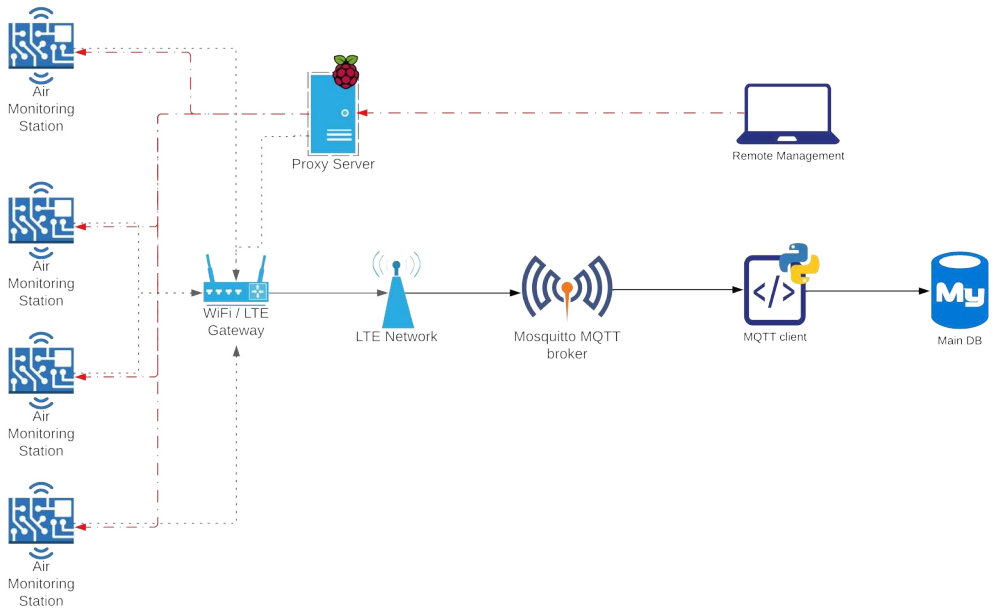


Fig. 4.6 Calibration Deployment

Urban Monitoring Using LoRaWAN

In this scenario, the performance of the LoRaWAN network in the metropolitan area of Turin and the use of an open network (The Things Network - TTN [193]) are evaluated. TTN provides LoRaWAN service through an open infrastructure and its subsequent cloud management. The evaluation considers parameters such as range and transmission optimization to obtain measurements at the highest possible resolution. Latency tests and more detailed examinations are presented in [194].

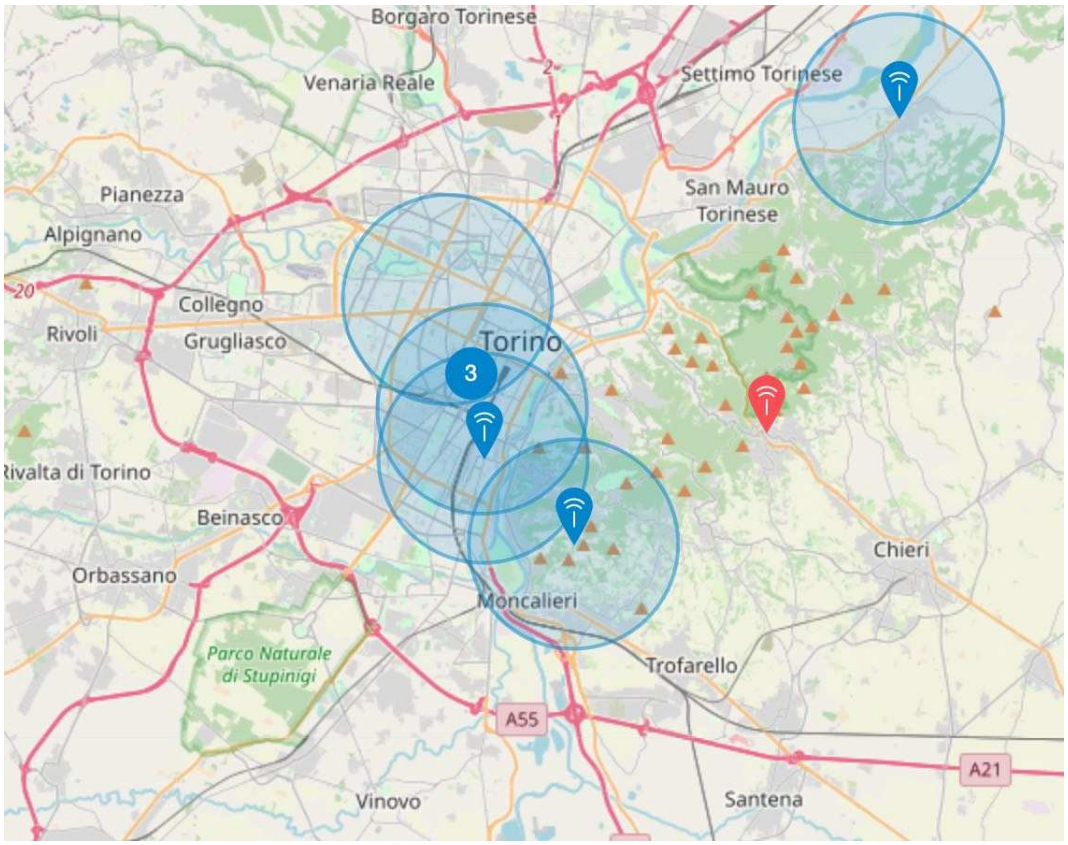
The metropolitan area of Turin has five gateways, as shown in Figure 4.7a. Although it covers a significant part of the urban area, this is an approximation. To determine the performance, transmission tests are conducted from four different geographical points and evaluated using a reference gateway (Polito). the points were selected to showcase different transmission environments, with the aim of emulating Line of Sight (LOS), Non-Line of Sight (NLOS), relatively short distances, and long distances (see Figure 4.7b. Point 1 was chosen to represent a far LOS scenario from the "Polito" gateway, featuring a few buildings of height less than 100 meters that did not obstruct the line of sight. Point 2 exemplified a short-range LOS scenario with a limited number of nearby buildings. Point 3 exhibited an environment with numerous obstructing buildings, impeding the view of the gateway. Lastly, Point 4 represented a distant location facing a water surface (river), where moving boats and various obstacles were present. The selected test points are enumerated in Table 4.2

Point	Distance from Reference Gateway (km)	Latitude (degrees)	Longitude (degrees)
Gateway	0	45.064167	7.6597
1	2.88 LOS	45.039404	7.645806
2	0.6 LOS	45.057617	7.659197
3	1.11 NLOS	45.061521	7.648052
4	2.59 NLOS	45.060447	7.694892

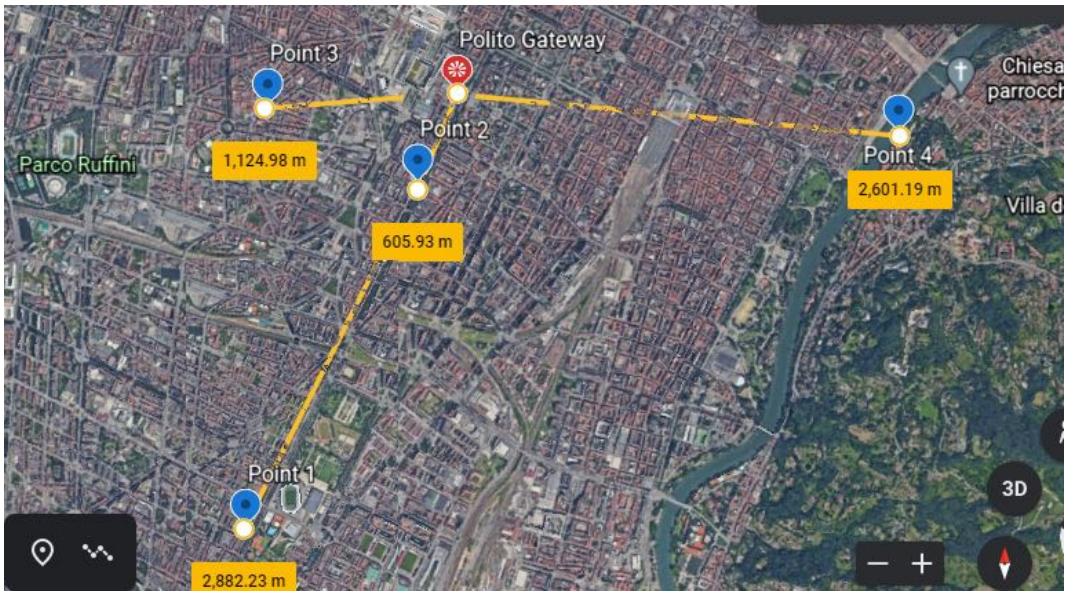
Table 4.2 LoRAWAN Testpoints

4.5.2 Participatory Sensor Networks

Participatory networks offer significant benefits for air quality monitoring systems. By leveraging strategically placed sensor stations and user smartphones, this approach allows for the collection and transmission of data. Mobile applications serve



(a)



(b)

Fig. 4.7 TTN LoRaWAN gateways: Ideal range coverage (a) and test points location from the reference gateway (b).

as a means of facilitating user interaction with the monitoring stations, enabling the gathering, adjustment, and transmission of data to an online collection point for processing and user access.

The integration of participatory sensing techniques in air pollution monitoring brings about numerous advantages. Firstly, it proves to be cost-effective as participants share their network data plans for transmitting the measurements. Secondly, the close proximity of the sensors to the environments where citizens reside enhances the value of the collected data. Moreover, involving users in the sensing process fosters citizen engagement and empowers individuals to actively contribute to environmental monitoring efforts. This active participation by users transforms the monitoring process into a more inclusive and participatory endeavor. Additionally, the incorporation of mobile applications within participatory sensing systems facilitates real-time data transmission and visualization. Users can access up-to-date information on air quality, enabling them to make informed decisions and take necessary precautions during periods of high pollution levels.

Nevertheless, this approach also presents challenges. Since data transmission from the microcontroller to the remote server relies on users, there is no guarantee of continuous data transmission. Privacy concerns may also arise as users may be hesitant to make their location information publicly available. Furthermore, the effectiveness of data transmission is contingent upon user interest, meaning that if few users participate, the measurements may not be frequently updated, resulting in reduced resolution.

In this scenario, the deployment of devices at fixed points of high attendance is proposed to increase the potential user base for this tool. In the deployment evaluated in this study, the architecture described in section 4.2 is used, adapting the mobile application proposed by [195] and modifying it for compatibility with the monitoring station presented in section 4.3.

The mobile application was developed using the Flutter framework, which enables the creation of a single codebase that can be deployed on various platforms such as Android and iOS. The application's core functionalities are integrated into different user interfaces within the app (see Figure 4.8. The authentication interface provides multiple login options, including traditional email and password login, as well as login through Facebook and Google accounts. This ensures a user-friendly and accessible login experience for participants. The home interface serves as the

main dashboard for data visualization. It utilizes visual components such as charts, graphs, and maps to present information based on the user's location. The concentration levels of PM_{10} and $PM_{2.5}$ are used to calculate the air quality index (AQI) using the NowCast algorithm [196], providing users with a clear perception of the current air quality. The map interface displays the user's current location and the locations of monitoring devices positioned throughout the city.

Another major challenge posed by participatory systems is data integrity. Users could manipulate and upload information to the system, generating readings that do not reflect the actual conditions. This deployment proposal to prevent such issues, the transmitted data is digitally signed by the monitoring station. The information is then validated by both the application and the server, which verifies the measurements before entering them into the database.

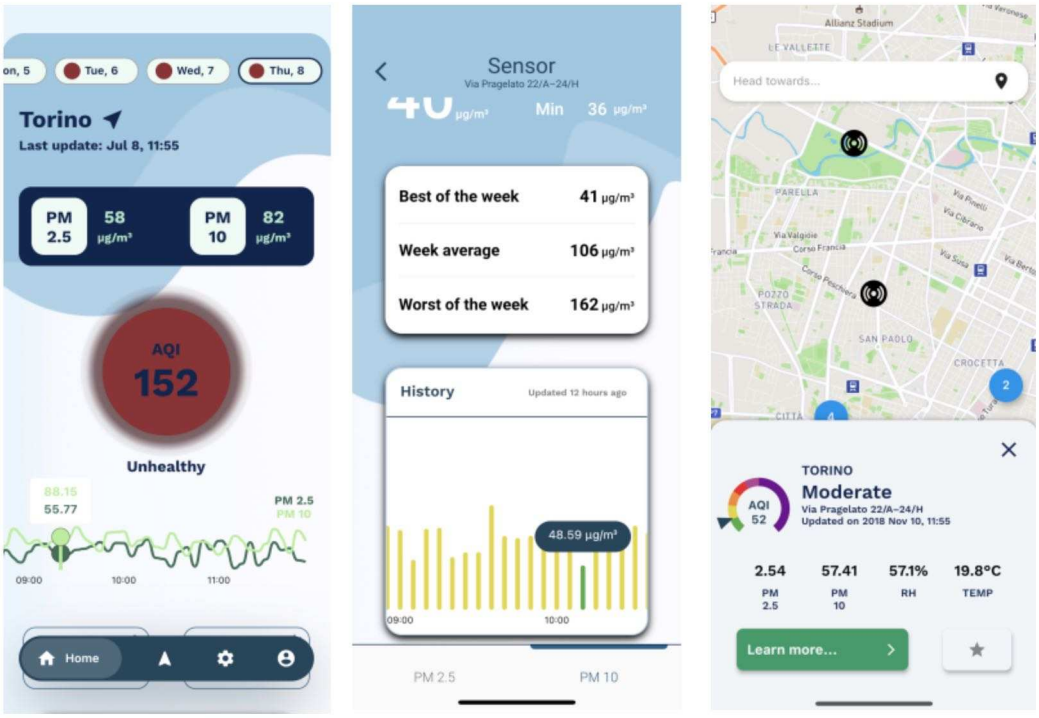


Fig. 4.8 Application screenshots

The evaluation of this architecture is conducted in a controlled environment, analyzing the transmission throughput as well as the sensor's performance in data transmission and integrity. A more detailed assessment of the entire system is presented in [197].

4.5.3 Mobile Sensor Networks

Mobile networks in smart cities typically rely on either mobile-based systems or systems that utilize crowd sensing. Systems utilizing vehicles often utilize the available public transportation in the city, such as buses, trams, and taxis.

In these mobile sensor networks, the main objective is to address the issue of coverage. A single sensor in such networks can monitor a larger geographical area, which is its main advantage. Depending on the speed at which the device is transported, large areas can be covered in a short amount of time. However, while the spatial resolution increases, the temporal resolution decreases as it depends on the frequency at which the sensor returns to the same area. Another advantage is maintenance, as sensors can be carried by users or transported to a maintenance point, reducing costs and facilitating tasks such as calibration.

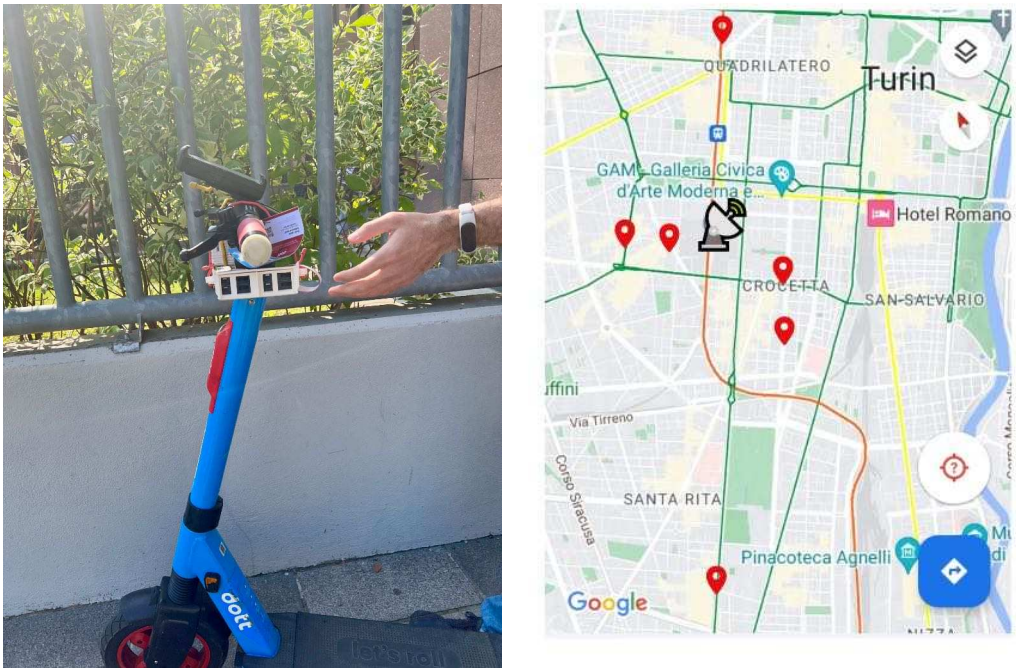
On the other hand, these networks also present challenges, such as the absolute dependence on GPS. GPS serves as the foundation of the system, but it demands more power and requires time to obtain a valid positioning value during startup. Highly urban areas or locations with signal obstacles can pose problems for monitoring stations. Furthermore, mobility is semi-controlled or uncontrolled, as it depends on public transportation routes or users, which can affect the resolution in less crowded areas or areas that are never visited. Conversely, it can also generate redundant data that does not provide additional information.

Regarding communications, using vehicles would require a constantly available communication system for each sensor, which increases costs. If a crowd sensing system is chosen, users share their data plans for transmission, eliminating costs but making the deployment dependent on the level of user involvement.

In this type of mobile network, this research proposes a deployment using a system based on public transportation and a crowd sensing system.

Urban Monitoring Using Electric Scooters and LoRa

In this scenario, the use of electric scooters as vehicles for attaching the monitoring station is proposed (see Figure 4.9a. Unlike public transportation, these vehicles can reach more distant locations beyond predefined routes, enabling greater coverage. Additionally, electric scooters are not affected by traffic congestion and their speed



(a) (b)

Fig. 4.9 Urban Monitoring Using Electric Scooter: Monitoring Station Placement on Scooter (a), and test points location from the reference gateway (b).

is limited to 20 km/h in Italy and 6 km/h in pedestrian areas [198]. However, they require WAN networks for data transmission, and the use of mobile networks incurs additional costs. Therefore, LoRaWAN is being evaluated as an alternative technology.

Another limitation in the measurement process is the lack of reliability of PM sensors in the presence of wind changes. Researchers such as [98] have found that the precision of PM measurements is affected by these fluctuations. As a result, PM measurements are only considered valid when the scooter is stationary.

Due to the variable coverage, the Adaptative Data Rate (ADR) algorithm is used in this deployment to optimize the radio link and improve the PDR of LoRaWAN. This mechanism automatically adjusts data transfer parameters and power based on the distance to the gateway, enhancing the performance of the system. For the evaluation, diverse locations within the city are selected using the same criteria as described in Section 4.5.1 to simulate different scenarios. Six locations were selected to represent LOS and NLOS conditions at varying distances (see Figure 4.9b). These

points and their distances are presented in Table 4.3. More detailed LoRaWAN ADR transmission experiments are presented in [199]

Location	Location Type	Distance (m)
Loc1	LOS	900
Loc2	LOS	1300
Loc3	LOS	2210
Loc4	NLOS	646
Loc5	NLOS	1300
Loc6	NLOS	2600

Table 4.3 LoRaWAN Test Points.

Crowd Sensing Deployment

The last scenario presents a crowd sensing scenario where the user acts as a volunteer in transporting the monitoring station. The user uses its mobile devices as a hotspot to facilitate the transmission of data generated by the monitoring station, without the need for mobile applications, thus protecting user privacy.

One of the challenges of this type of deployment is the resilience of the data, as it may be collected from within vehicles, which may not reflect the true level of PM in the area. Additionally, multiple users may be present at the same location, generating redundant data. To address these issues, it is suggested to use this approach for specific studies that require detailed analysis of a specific area.

In this research, monitoring is proposed at specific points through coordinated users. An experiment is conducted to monitor and locate areas with more pollution levels in a neighborhood, following predetermined monitoring routes within a specific time range. This study is conducted in the Milan Barrier neighborhood (45° 05'N 7°42'E) in the city of Turin, as a continuation of the case presented in Chapter 2.2.

4.6 Results and Discussion

Calibration Scenario

In this scenario, ten monitoring stations were deployed on the roof of the Rubino ARPA station, as shown in Figure 4.10.

Fig. 4.10 Air Monitoring Station Deployment on Calibration Stage

Each stored measurement, there are three primary components of information: the timestamp when the measurement was taken, the identifier of the sensor that collected the data, and the actual value of the measurement. The encoding of each measurement is performed in text format, and each MQTT message is transmitted every minute, following a structure similar to the example provided below:

```
...
2022-09-19 12:30:05,49,12
2022-09-19 12:30:06,129,35.6
2022-09-19 12:30:07,15,25.8564
2022-09-19 12:30:08,46,9956.45
2022-05-24 09:03:29,60,45.06411
2022-05-24 09:03:29,60,7.659407
...
```

In this experiment, the sampling intervals were configured as follows: 1 second for PM sensing, 2 seconds for temperature and humidity, 5 seconds for atmospheric pressure, and location updates were obtained every minute. Each MQTT message had an average payload size of 13 KB, resulting in an average daily transmission of 18.7 MB per device. Consequently, the total monthly transmission across all stations amounted to 5.7 GB.

While this level of traffic generation is generally sufficient for an average data plan of 5 GB per month, it posed challenges for the microcontroller's operations.

Both the I/O operations and wireless network transmission were incompatible with the larger file size. The transmission and I/O process required more than 1 second of computations in some cases, leading to potential data loss from the sensors during this time due to the Micropython's interruption handle. Consequently, the combined impact of I/O operations and message size could compromise the integrity of data transmission.

To enhance the transmission, the data was converted into binary code using a big-endian representation. Each piece of information, including the timestamp, sensor identifier, and measurement values, was converted into bytes. The timestamp was first converted from the ISO 8601 format to the UNIX format and then represented as a 4-byte signed integer. The sensor identifier was converted into a 2-byte short integer, one-byte to indicate the type of sensor, while the measurement values were treated as 4-byte floats for atmospheric pressure and GPS, meanwhile for temperature, humidity and PM a 2-byte short integer is used. The resulting binary representation of the data, with bytes displayed in hexadecimal format, is shown in Figure 4.11.

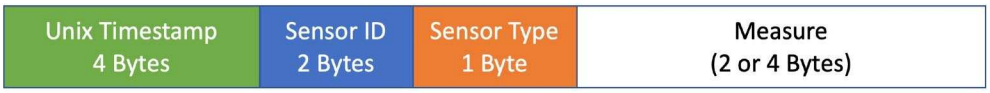


Fig. 4.11 Example of the binary format

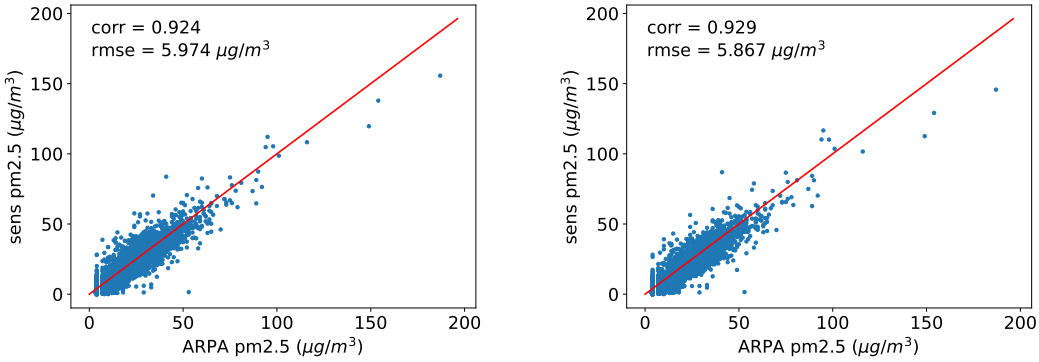
Under this data encoding, the size of the MQTT message payload is reduced to 5KB, consequently diminishing the amount of daily data generated to merely 7.2MB. This results in a significant reduction of 61.5%, thereby limiting measurement losses in other processes to a negligible 0.22% of daily measurements. Regarding transmission efficiency, the PDR remained unchanged due to the stability of the connection throughout the evaluated period, with the only losses occurring during maintenance windows.

In this deployment scenario, the performance of five monitoring stations in measuring particulates was examined over the period from October 2022 to April 2023. The reference for this evaluation was based on the readings obtained from the ARPA station. The average Root Mean Square Error (RMSE) for each station, along with their respective Pearson's correlation coefficients, are summarized in Table 4.4

In this analytical evaluation, the station's performance demonstrated a strong correlation when compared to benchmark values sourced from existing literature.

Board ID	**RMSE**	r^2
Board 20	7,1075	0,87325
Board 21	7,9335	0,87175
Board 25	6,578	0,9255
Board 29	7,27375	0,90225
Board 31	6,38375	0,924

Table 4.4 Evaluation of the designed monitor station's data alignment with ARPA's reference equipment using Pearson's Coefficient (r^2) and RMSE



(a) Sensor #269 - Board 21.

(b) Sensor #393 - Board 31.

Fig. 4.12 Scatter Plots for Air Monitor Sensors. Board 21 (a), Board 31 (b).

Notably, upon manual inspection a sensor deployed on Board 21 was identified to have underperformed due to a malfunction. Despite this anomaly, the Particulate Matter (PM) sensors deployed across the board showed a consistent pattern of strong correlation and alignment with the measurements taken by the benchmark station.

This consistency was further illustrated in Figure 4.12, which visually represents the dispersion results of the PM sensors. These graphs effectively underscore the significant correlation that exists between the ARPA data and the data captured by a PM sensor housed within our custom-designed monitoring stations.

In an effort to further examine the operational sensors an in-depth analysis is executed, which revealed compelling results. The data exhibited an average correlation coefficient that was approximately 0.9. Additionally, the RMSE showed a mean value of 6.72. Intriguingly, these figures surpass the calibrated values presented in studies like [13].

This analysis is pivotal as it not only affirms the robustness and reliability of the monitoring stations in the proposed deployments, but it also emphasizes their

potential to generate accurate, reliable data. This inherent capacity for precision, as demonstrated by the higher than expected values, assures a high level of fidelity in the data collected, thereby enhancing the overall quality and effectiveness of the research.

Urban Monitoring Using LoRaWAN

In this deployment, the encoding of the message differs from the previous scenario due to the limited throughput, necessitating a single message to transmit the data from all sensors. For this purpose, a unique message is proposed, encompassing the $PM_{2.5}$ and PM_{10} values. Each sensor requires 10 bits, which translates into a need for 5 bytes per particle size for the four sensors. A byte is used for temperature and one byte for the humidity, 4 bytes for GPS data, and another 4 bytes for the timestamp of the message transmission. This message is shown in Figure 4.13.

Fig. 4.13 Message packet format for the LoRa transmission

The sampling periods were set for this scenario as follows: 1 second for PM sensing, 2 seconds for temperature and humidity, location updates were obtained every minute and the atmospheric pressure was disabled.

Each message transmits the average measurements taken between each message, adhering to the transmission duty cycle defined for LoRa, which is dependent on the utilized Data Rate (DR) in LoRa. To evaluate the range, a fixed payload data of 33 bytes (comprising 20 bytes of payload and 13 bytes from headers) will be transmitted with varying DR. These DRs are determined by the Spreading Factor (SF) and the Bandwidth. For The Things Network (TTN), the recommended bandwidth is 125 kHz, with SF values ranging from 7 to 12 (equivalent to DR5 to DR0; refer to Table 4.5). A total of 100 (unconfirmed) packets will be transmitted from the sensor node to the Gateway.

Figure 4.14 illustrates the results of the PDR for each test point, with the spreading factor varied for the proposed payload. An almost 100% success rate is achieved

Data Rate	SF	BW (KHz)	CR	Bit Rate(bps)
DR5	7	125	4/5	5468
DR4	8	125	4/5	3125
DR3	9	125	4/5	1775
DR2	10	125	4/5	976
DR1	11	125	4/5	537
DR0	12	125	4/5	292

Table 4.5 LoRa Data Rates (DR) parameters for the EU868 region

at all distances with SF12, whereas SF7 and SF9 exhibited unstable behaviors contrary to expectations. Improved performance for the same Spreading Factors (7 and 9) was observed. Additionally, as the distance increases (as in Test Points TP3, TP4), a clear decline in the PDR for this payload can be identified, which aligns with the physical description of the SF.

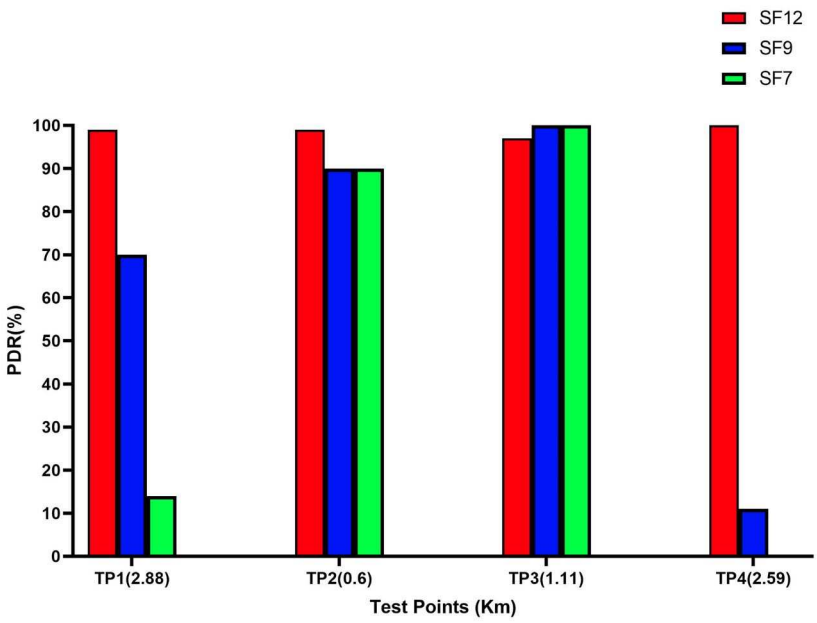


Fig. 4.14 PDR for the proposed payload

It's important to note that these PDR measurements were taken solely with respect to the "Polito" Gateway. Nevertheless, the TTN web server indicated a much higher PDR for packet reception through other Gateway. However, that was disregarded due to their erroneous positioning (more than 20 km), rendering it unreliable. Furthermore, an elevation in the code rate can lead to improvements in the range performance. The minimum Code Rate (CR) of 4/5 was used in our

work, but employing a CR of 4/8 could result in a broader range (better reception for the same Node). However, this advantage is offset by a corresponding reduction in bit-rate.

Continuing from the previous discussion, the ensuing paragraphs delves into an analysis of the applicability of data rates that LoRaWAN supports for the system in question. This evaluation incorporates the limitations of the duty cycle and TTN fair access policy, assessing whether the technology can deliver packets at a rate deemed reasonable. The analysis is initiated with a straightforward examination of the time required to transmit a 20-byte payload packet, adhering to all imposed restrictions. Utilizing the airtime calculator [200], a satisfactory approximation of the on-air Time-On-Air (TOA) of the packet can be ascertained, along with the subsequent duty cycle, device sleep duration, and the number of packets (messages) that can be dispatched within an hour. Thus, the total packet size amounts to 33 bytes (20 + 13). Table 4.6 outlines the timing analysis for each Data Rate (DR).

DR	**TOA(ms)**	**DC 1%(sec)**	**Sleep Time(min.)**	**Max. MSG/Day**
DR5	71.9	7.2	3.5	417
DR4	133.6	13.4	6.4	224
DR3	246.8	24.7	11.8	121
DR2	452.6	45.3	21.7	66
DR1	987.1	98.7	47.4	30
DR0	1810.4	181	86.9	16

Table 4.6 TTN Timing Analysis for 20 bytes payload

With the bit rate information amalgamated with LoRaWAN transmission restrictions and the maximum packet rate of the weather station not exceeding 5 minutes, it becomes evident that the most suitable mode for transmission for this system is SF7. It is important to note that the TTN Fair Access Policy is not applicable to private web servers. However, if the deployment utilizes the open network, the system operates under the constraints imposed by the TTN server, translating to a bit rate of 5.5 Kbps and a T_{off} of 3 minutes. It is noteworthy that the T_{off} fluctuates in accordance with the size of the packet.

These findings suggest that each message transmits the average measurements taken every 3.5 minutes (approximately 207 seconds). This applies under ideal distance conditions. For more remote points, where the PDR is insufficient for a high DR, the messages will convey larger averages, approaching a minimum of 16

messages per day in the worst-case scenario (DR0). To alleviate these constraints, a modification of the message encoding is proposed.

The format for this is presented in Figure 4.15. As the stations are stationary, the transmission of GPS data is unnecessary, thus freeing up this space. Furthermore, owing to the redundancy of sensors, which can exhibit failures and noise, only the transmission of one particulate value is necessary. Therefore, only the median of the averages of each sensor between messages is transmitted, requiring just 10 bits per type of particulate. With regard to time, a timestamp is employed to determine the measurement time, indicating the number of seconds in the current day. This method necessitates only 9 bits for transmission.

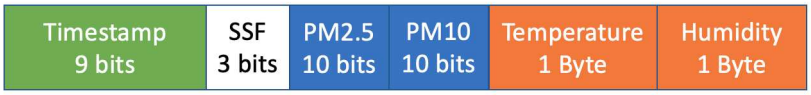


Fig. 4.15 Reduced Message format with sensor states flags SSF

As weather data does not fluctuate rapidly (on a per-second basis) and values may remain consistent for certain periods of the day, it is not necessary to transmit data at the same frequency at which it is produced by the sensors. Instead, a simpler approach, such as the introduction of Sensor States Flags (SSF), can be utilized. The SSF serves as a bit set that indicates any changes in the sensor readings and can be used to identify the measurements included in the message. This results in the message now having a variable length, ranging from 6 to 3 bytes. The coding for these flags is presented in Table 4.7.

PM bit bit 2	**T bit bit 1**	**H bit bit 0**	**Packet Measures**	**Packet Size**
0	0	0	Not Valid	-
0	0	1	H	3
0	1	0	T	3
0	1	1	T + H	4
1	0	0	PM	4
1	0	1	PM + H	5
1	1	0	PM + T	5
1	1	1	All Measurements	6

Table 4.7 Encoding of the SSF

The same test was repeated using this condensed format, with the results displayed in Figure 4.16. Once again, SF12 achieved nearly 100% success across all distances for shorter payloads, yet exhibited improved PDR compared to the longer message. Unexpectedly, SF7 and SF9 displayed instability, despite predictions to the contrary. It was anticipated that Test Point 2 (TP2), being closest to our Gateway, would present the highest PDR for shorter payloads. However, the results revealed superior performance at Test Point 3 (TP3) for these DRs. This anomaly may be attributed to various forms of interference that affected the LoRa device during the experiment. The presence of vehicles, bus movements, and potentially other radio signals at TP2, which is a bus station, could have contributed to this behavior.

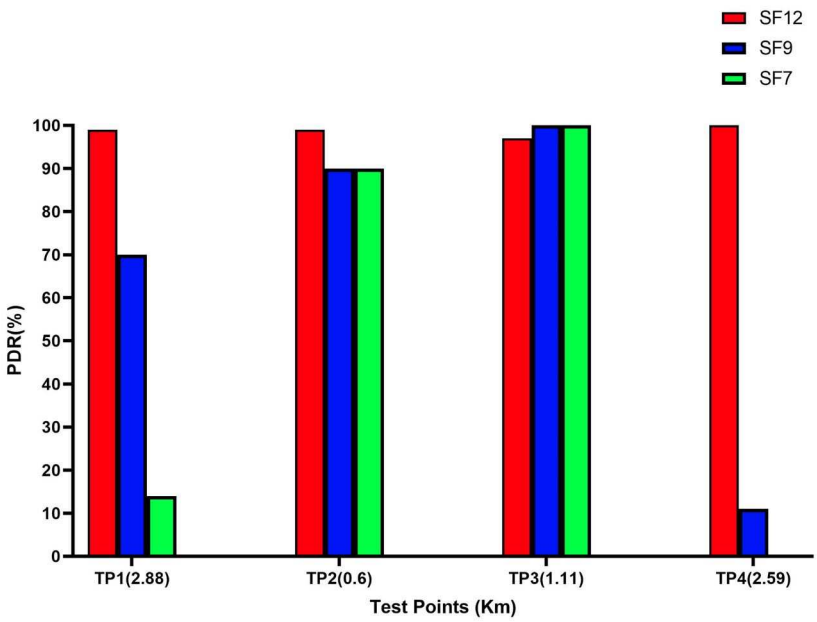


Fig. 4.16 PDR for the shorter payload

As the message is variable, the number of messages changes depending on the minimum (3 bytes) and maximum (6 bytes) message sizes. Using the airtime calculator again, the minimum time is reduced to 148 seconds (2.5 minutes) between messages, which is a one-minute decrease from the previous version (see Table 4.8. These values are dramatically affected by the TTN Fair Access Policy, whereas under a private network, the values would be closer to a 5-second interval between messages, complying with the sampling frequencies presented in Chapter 3. Therefore, under an open LoRaWAN network, substantial limitations are generated, making this

	Minimum Message 3 Bytes (16 with header)		Maximum Message 6 Bytes (19 with header)	
DR	**Sleep Time (min)**	**Max Msg/Day**	**Sleep Time (min)**	**Max Msg/Day**
DR5	2.5	583	2.5	583
DR4	4.4	323	4.9	291
DR3	7.9	181	8.9	161
DR2	15.8	90	15.8	90
DR1	31.7	45	35.6	40
DR0	63.3	22	63.3	22

Table 4.8 TTN Timing Analysis for 3 and 6 bytes payload

deployment suitable mainly for monitoring areas with slow changes in particulate concentration, such as parks or areas with distant pollution sources.

Participatory Sensor Network Deployment

In this scenario, where BLE is employed, the transfer rates are constrained, thus requiring optimization of the transmission process. The format described in Figure 4.11 is employed again for this purpose, facilitating the generation of files containing the measurements. For this implementation, the measuring file stored in the SD changes from one-day measurements file, to a one-minute measurements file. To ensure data integrity, a digital signature is incorporated into the measurement files. Utilizing asymmetric cryptography, this digital signature verifies the sender's identity and confirms the message content's inviolability during transmission. The process involves hashing the message with an SHA256 algorithm, then encrypting it using the sender's private RSA key. The recipient validates the signature by hashing the message, decrypting the signature using the sender's public key, and comparing the two hashes. The byte array containing the measurements is converted into a string via Base64 encoding to facilitate this.

Despite its efficacy, digital signature creation is computationally expensive and could lead to data loss during the two seconds required for its formation. Moreover, the high computational cost along with I/O operations could affect data transmission, particularly because transmission primarily occurs when a smartphone is nearby, requiring prioritization of this operation. To mitigate this, a dedicated module is introduced between the business logic and the file system. This module manages a buffer queue system prioritizing different operations, enhancing synchronization

and performance, and reducing frequent SD memory access (see Figure 4.17). This system stores copies of some measurements and signature files from the file system in RAM buffers. The module comprises:

- A queue of the most recent files, both measurements and signatures, which are readily available for transmission when a smartphone connects to the device.
- A queue of the latest measurement files ready to be signed. This computationally expensive process is managed via a priority queue that is processed in one-minute intervals, provided the device is not involved in transmission.
- A queue of files already transmitted via BLE, waiting for receipt confirmation.
- A buffer containing data for the current minute's temporal window. Upon changing the minute, this data is saved in a file system and moved to the top of the signing queue. This buffer decouples the sensing module's logic from the file system management.

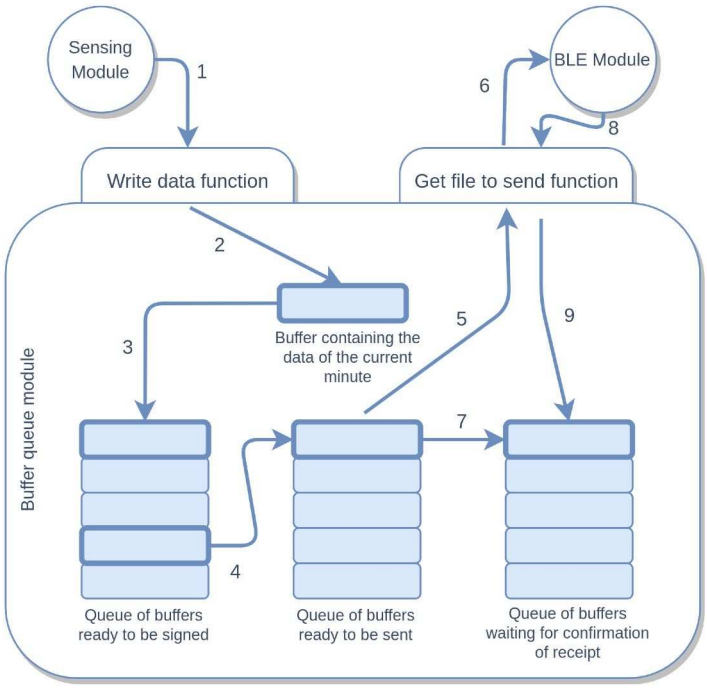


Fig. 4.17 Buffer queue system

The application on the smartphone, serving as the GATT Server, receives data from the monitoring device, acting as a GATT Client, using Bluetooth Low Energy

(BLE) technology and the ATT protocol. This connection is established on a one-on-one basis.

Transmission of the files considers the limitations set by the MTU, which caps the channel size to 185 bytes due to mobile operating system constraints. This necessitates the partitioning of files into 182-byte chunks to account for the 1-byte ATT header and 2-byte ATT handle that occupy 3 bytes of each ATT packet. This method enables the transmission of a file representing one minute of measurements in less than 500 ms, significantly faster than the 4-5 seconds required to transmit equivalent CSV data.

Reducing file size enhances transmission reliability by lowering the risk of transmission errors due to smaller byte quantity. Additionally, the decision to limit the temporal window of files to one-minute file was influenced by multiple factors: reducing potential data loss from transmission errors, ensuring data relevance for corrective action due to the variability of particulate matter concentration, and accommodating the limited proximity of users to the station, as BLE operates effectively only within a 10-meter range.

Challenges arose during the buffer queue system development when the file system was initially configured to house all measurement files in a single folder. This led to I/O operation times soaring from milliseconds to seconds after a few days of data collection as the number of files reached into the thousands. To mitigate this, the file system was restructured into a tree-like format. In this new structure, root folders, named by the date of measurements, contained up to 24 subfolders representing each hour of the day. Each of these hourly subfolders could hold a maximum of 60 files, each representing a minute within that hour. This solution, while adding complexity to the logic and file system, curtailed the read/write operation time to less than 150 ms per file.

Table 4.9 presents a selection of the results gathered. The evaluations were performed using both textual (first column) and binary (second column) measurement files, each encapsulating data over a one-minute temporal window and containing an equivalent number of measurements. The figures in the table represent the mean values for an individual file, calculated as the average of all test outcomes.

Given that the signature generation process necessitates a 2-second computation period during which the device is unable to manage interrupts from the sensor timers, this leads to data loss each minute as the interrupts can only be processed when the

	Text files	Binary Files	Improvement(%)
File size	~14.5KB	~5 KB	~65.76 %
I/O operations time	$\geq$ 500 ms	~150 ms	$\geq$ 70 %
Data transmission time	~4500 ms	~500 ms	88 %

Table 4.9 Data storage and I/O operations statistics

thread is available. Moreover, as each sensor follows a unique data sampling regimen, the degree of data loss fluctuates, yielding a range from minimum to maximum data loss percentages. The table 4.10 delineates the sampling period for each sensor and the number of measurements included in each sample, allowing for computation of these data loss values:

Sensor type	Sampling time	N° of easurements per sample
PM_{10} and $PM_{2.5}$	1 s	$2*4=8$
GPS (latitude, longitude)	60 s	2
DHT (T, RH)	2 s	2
BPM (Atm Pressure)	5 s	1

Table 4.10 Sensors' data sampling periods and number of measurements

Given the variables of measurements within a sample (m_s), sampling time (t_s), and the count of each sensor type affixed to the board (n_s), A theoretical maximum quantity of measurements (M) accommodated in a singular file having a temporal window of one minute can be calculated by:

$$\begin{aligned} M &= \sum_s m_\mathrm{s} \frac{60}{t_\mathrm{s}} n_\mathrm{s} \\ &= 2\frac{60}{1}4 + 2\frac{60}{60}1 + 2\frac{60}{2}1 + 1\frac{60}{5}1 \\ &= 480 + 2 + 60 + 12 = \mathbf{554} \end{aligned} \tag{4.1}$$

Considering that the signature process suspends the capture of new measurements for about 2 seconds, PM sensors are the most impacted due to their sampling frequency, leading to a minimum loss of 8 measurements per minute. This value can increase to 21 measurements if all the timers coincide. Ultimately, the data loss percentage for each sensor can be determined under both optimal and least favorable conditions, revealing a data loss rate between **1.4%** and **3.8%** every minute.

Finally, several communication strategies were explored to facilitate the rapid transmission of data through BLE from the monitoring device to the mobile application. The optimal approach was empirically chosen based on performance metrics. Initially, multiple files were corrupted during the BLE transmission, making them unreadable by the client. Consequently, a signature process was implemented to enable verification of file integrity on the client's side. Despite the inevitable performance degradation, the server needed to ascertain the success of each transmission to facilitate the retransmission or deletion of specific files.

Performance was influenced by several variables. One was the mix of commands offered by Bluetooth Low Energy technology, such as *write*, *read*, and *notify/indicate*. Another crucial factor was the buffer queue system that aided the data exchange protocol by storing files in RAM, thereby minimizing I/O operations with the SD card during transmission.

Table 4.11 presents the results in terms of data transmission time for each measurement file and its corresponding signature, considering the various solution combinations. Depending on the chosen solution, data sending can require either read or notify operations. Two of these solutions pair these with a write operation, confirming file receipt, while others do not await receipt acknowledgment.

BLE commands	**Without buffer queue system**	**With buffer queue system**
Read and Write	~4900 ms	~4700 ms
Read	~1100 ms	~950 ms
Notify and Write	~650 ms	**~500 ms**
Notify	~450 ms	~300 ms

Table 4.11 Data transmission time statistics

The findings indicate that read operations are generally slower than notify ones and that acknowledging each chunk of the file is costly from a performance perspective. The optimal solution employs the notify approach paired with write operations for improved reliability, despite an approximate 200 ms increase in transmission time. Moreover, utilizing the buffer queue system as a memory access point for file caching in RAM reduces overall transmission time by around 150 ms, equivalent to the I/O time.

The results featured in the table were derived from experiments primarily conducted with devices placed within two meters of each other. Additional experiments to assess the influence of Bluetooth signal attenuation on transmission time did

not reveal significant variations. Specifically, the signal power is undetectable at a distance of roughly 10 meters, resulting in a disrupted connection.

The aforementioned results do not account for the time needed to establish a BLE connection between the devices. This time, independent of the communication command used and required only once at the onset, averaged around 2 seconds. This was measured in the app, starting from the board's discovery through Bluetooth scan until the receipt of the first chunk of a file. Adding this time to the time required to transmit all chunks of a file using notify and write commands indicates the minimum time a user should be in proximity to the board to enable a full transmission.

Given the times obtained in this deployment scenario, locations where users tend to stay for approximately a minute are necessary to ensure that the connection and transmission processes occur successfully. Constant user activity is also needed for frequent transmission of the measurement files. Public service stations serve as ideal environments for this type of deployment.

Urban Monitoring Using Electric Scooters and LoRa

This deployment displays characteristics similar to those described in the previous LoRa scenario, with the sensors operating at the same sampling periods. The main difference is the use of a private LoRaWAN network, which removes constraints from usage policies, enabling a higher number of message transmissions in each DR. This in turn allows for the transmission of higher-resolution information per message. The transmission format proposed in Figure 4.18, which consists of a 33-byte payload, is used for this purpose without losing granularity for each PM sensor and precision for GPS location and temperature.

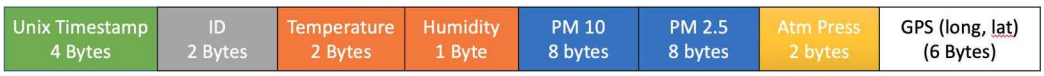


Fig. 4.18 Message packet format for the LoRa transmission

Sixty data packets were dispatched, with a 30-second interval between each message in each point. The status and performance of the gateway can be assessed via the analysis of the Received Signal Strength Indicator (RSSI) and Signal-to-Noise Ratio (SNR) values.

As the two key components in determining the quality of the received radio signal, RSSI and SNR are measured by the LoRa Gateway upon receipt of each communication. RSSI, the power of the received signal in milliwatts, is gauged in decibels (dBm). Its values span between -30 dBm and -120 dBm, with -30 dBm indicating a very strong signal and -120 dBm pointing to a very weak one. SNR, the ratio between the received signal power and the power level of the noise floor, ranges between -20 and +10 dB. A figure near +10 dB signifies that the received signal is less corrupted. It is notable that LoRa can demodulate signals from -7.5 dB to -20 dB below the noise floor.

TTN uses the 20 most recent uplinks from the moment that the Adaptive Data Rate (ADR) bit is set. These measurements, comprising the frame counter, SNR, and the count of gateways that received each uplink, are used to calculate the margin, the difference between the measured SNR and the required SNR to demodulate a message given the Data Rate (DR):

$$Margin = SNR - SNR_{demodultaion} \tag{4.2}$$

This margin indicates the potential for DR to be increased or the transmission power to be reduced, thereby optimizing the Time-on-Air (TOA).

In the assessment of this scenario, average RSSI, average SNR, and PDR were measured for each location as described in Table 4.3. Initially considering the LOS locations, all points recorded a positive SNR, enabling a high PDR. For locations 2 (1300m) and 3 (2210m), the SNR remains high (positive); however, the PDR begins to decrease, which triggers the activation of the ADR algorithm, reducing the DRD from DR5 to DR4 in both points.

In the case of the NLOS locations, similar results are obtained. The RSSI decreases due to the interference caused by obstacles. For instance, location 4 (646m) experiences interference due to tall buildings between the device and the gateway. However, the DR retains its highest value. For the last locations, 4 and 5, located 1300m and 2600m away respectively, the RSSI is affected by more obstacles, impacting both the SNR and the PDR. The ADR is adjusted for both locations 5 and 6 to DR4 and DR3, respectively. These results are summarized in Table 4.12.

Based on the transmission results obtained, it's possible to maintain the message frequency for the given message size, thanks to the ADR. However, the Packet

Location	Location Time/Distance	DR	RSSI (dbm)	SNR (dB)	PDR(%)
Loc1	LOS/430m	DR5	-109	7.8	100%
Loc2	LOS/876m	DR4	-109	6	90%
Loc3	LOS/2210m	DR4	-110	3.8	80%
Loc4	NLOS/646m	DR5	-101	4.8	95%
Loc5	NLOS/1700m	DR4	-109	5.8	76%
Loc6	NLOS/3000m	DR3	-114	0.5	73%

Table 4.12 Mobile LoRaWAN Performance

Delivery Ratio (PDR) tends to deteriorate, potentially causing losses in the resolution of the measurements. Smaller message formats allow for an increase in message frequency, which mitigates resolution loss. For instance, a format like the one presented in Figure 4.15 enables minimum message periods between 5.1 seconds and 18.5 seconds for the same range of DR. While this would provide higher resolution, it implies greater power consumption by the monitoring station, thereby limiting autonomy.

While the mobile scenario is not a target from the deployment standpoint, since the ADR is not optimized for mobile transmitters [201], it still provides valuable data. During the tests between points, data is recorded to determine if, given the speed limit of the electric scooter, it is possible to determine the path and pollution values generated, considering an ideal PDR and altering the GPS sampling period to 3 seconds. Figure 4.19a presents the trajectory recorded within the device at maximum frequency, contrasted with the values transmitted every 30 seconds (Fig. 4.19b).

From these trials, it becomes evident that the low speed of electric scooters combined with the transmission frequency is sufficient for the detection of specific points with higher pollution levels, as well as for determining the trajectory across the city.

Crowd Sensing Deployment

Lastly, this scenario is based on WiFi transmission between the monitoring station and a mobile device acting as a WiFi hotspot. Under this modality, the transmission of measurements will depend on the cellular coverage in the area of interest, as WiFi

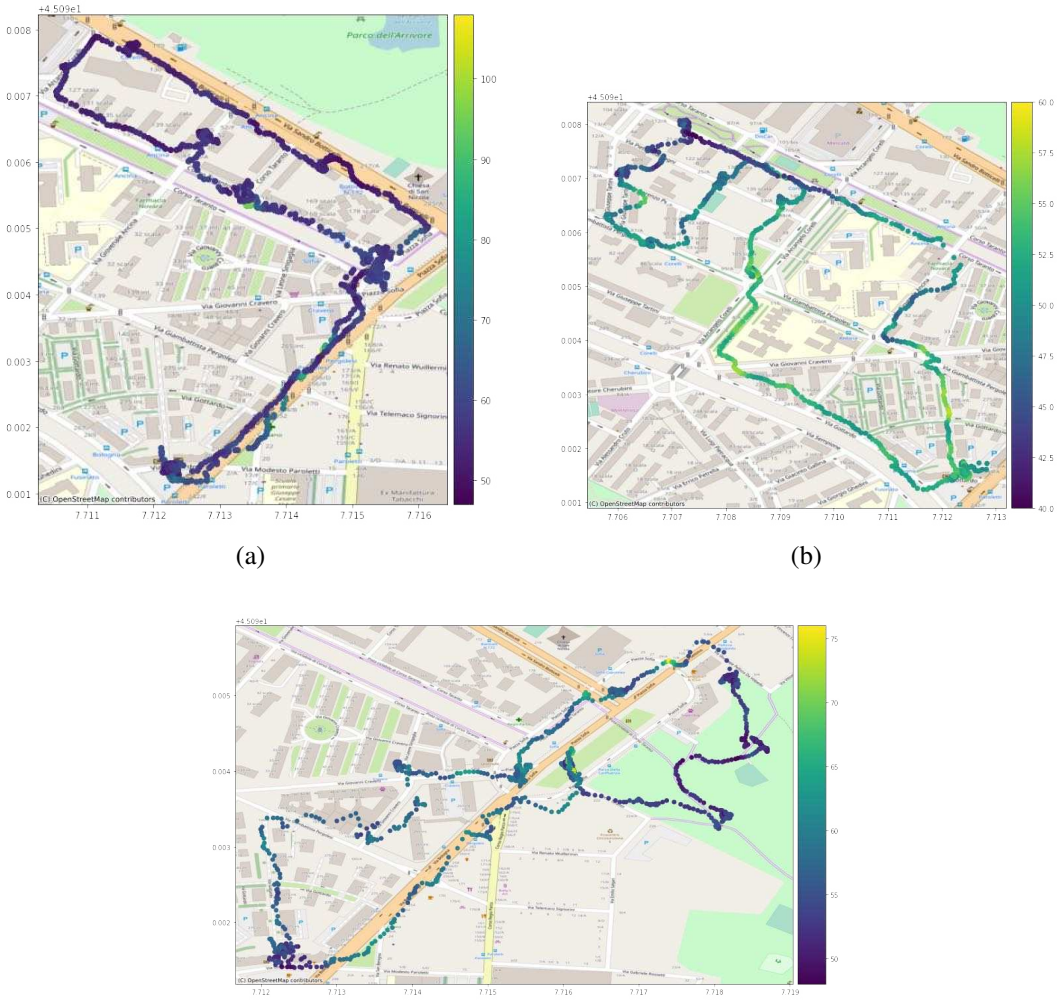
Fig. 4.19 Urban Rides Using Electric Scooter: Complete measured data at the monitoring station (a) and data transmitted every 30 seconds (b).

communication under normal conditions is considered reliable due to the proximity between the mobile device and the monitoring station.

This deployment was also evaluated within various routes in the Milan Barrier neighborhood. Here, several volunteers performed a route gathering pollution data to detect areas of higher exposure. Some routes are illustrated in Figure 4.20.

The execution of these routes with this deployment offers the capacity to produce intricately detailed maps focused on distinct areas. Specifically in this case, the approach allows for the intensive monitoring of a particular neighborhood, characterized by its high spatio-temporal resolution. This level of detail does not only illustrate a more precise pollution landscape but also fosters a more comprehensive understanding of the environmental dynamics within the area.

The generated information proves itself to be considerably valuable, especially when conducting specialized architectural analyses. These analyses can particularly revolve around topics related to sustainability, providing insights on how to design or modify structures and systems to enhance their environmental compatibility. An in-depth exploration of this utility has been discussed previously in Chapter



(a) (b)

(c)

Fig. 4.20 Crowd Sensing routes in the Milano Barrier neighborhood within a two-hour window. North (a), West (b), South and East (c).

2.2. By leveraging this approach, researchers and urban planners can identify pollution hotspots and patterns within a neighborhood over time. In turn, this can guide sustainable urban planning and policy-making, helping to create healthier and more sustainable living environments. The fusion of detailed monitoring data and sustainability-oriented architectural analysis thus offers promising possibilities for both environmental research and urban development.

4.7 Conclusions

The potential transformative role of the Internet of Things (IoT) in creating smart cities is a widely recognized fact. However, the success of such an application, or the insightful analysis of the targeted phenomena, hinges critically on the strategic deployment model employed. Achieving robust spatiotemporal resolution of certain variables necessitates a delicate balance between operationally viable solutions and the provision of valuable, actionable information.

Various deployment methodologies and network topologies have been the focus of extensive research. This research has been instrumental in defining a complex interplay between three pivotal objectives: cost-effectiveness, temporal resolution, and spatial resolution. Each facet of this triadic framework is targeted for maximization, but due to inherent limitations of each data transmission technology and the specifics of the deployment scenario, simultaneous maximization of these objectives remains elusive.

This study conducted an in-depth analysis of the three main deployment scenarios commonly utilized in the context of smart cities, namely fixed sensor networks, participatory sensor networks, and mobile sensor networks. While each of these deployment scenarios offers unique advantages, none presents a comprehensive solution when employed in isolation. Hence, this study introduces a novel hybrid model that integrates these three distinct scenarios, employing a single architecture and a standardized particulate matter monitoring station.

The system architecture proposed here is noteworthy for the use of wireless technologies working on unlicensed frequency bands, such as LoRa, WiFi, and Bluetooth. These technologies' penetration into the consumer market ensures the possibility of broad participation in data collection, thus contributing to the dual objectives of cost-

effectiveness and high spatiotemporal resolutions. The architecture's simplicity also ensures ease of data centralization and management. The reliance on straightforward, scalable data exchange protocols like MQTT and REST facilitates ease of integration and scalability. Furthermore, this model's inherent flexibility, which allows a single station to interface with a range of communication technologies prevalent in large urban areas, presents a significant advantage for future IoT deployments.

The findings of this study carry far-reaching implications, suggesting new pathways for harnessing the potential of IoT in urban environments, shaping smart cities, and deepening our understanding of complex urban phenomena through sophisticated data collection and analysis strategies. This real-time data availability also benefits urban planners, policymakers, and researchers who can utilize the collected data for urban planning, policy formulation, and scientific analysis. In turn, this can help in creating healthier and more sustainable living environments. The fusion of detailed monitoring data and sustainability-oriented architectural analysis thus offers promising possibilities for both environmental research and urban development.

Chapter 5

Data Accuracy

Air quality tracking and PM particle monitoring traditionally rely on sparse networks of stationary, reference-grade detectors, as dictated by European regulations 50/2008/CE and 107/2004/CE. The spatial coverage of these networks is constrained due to the high cost of instrumentation, with micro-balance PM monitoring stations costing up to 100K dollars [202]. Furthermore, these sparse networks might not adequately capture the substantial spatial variability often exhibited by air pollutant concentrations.

In response to these challenges, the development and application of affordable, portable, light-scattering-based PM detectors, costing merely a few hundred dollars [202], has surged recently. These cost-effective sensing nodes find wide applications [203–213], ranging from city-wide deployments, bike sharing fleets, location-aware places, and moving vehicles, to the monitoring of historic building conservation states and forest fire detection. They even enable participatory sensing through mobile monitoring.

However, these mobile sensing devices come with their share of challenges, including the limited processing capabilities, and concerns over precision and high variance. These drawbacks necessitate an adequate engineering approach to be overcome effectively.

The identification of anomalies forms a critical facet in the realm of sensor networks. This involves detecting unexpected sensor behaviors that lead to erroneous readings, either in single measures or data sequences. As sensor malfunctions and failures can significantly impact the calibration process, which relies on precise

and consistent measurements for generating an accurate fitting model, this aspect requires serious consideration.

Therefore, calibration plays a pivotal role in determining the final accuracy and reliability of a sensor network. To mitigate the impact of the most common sensor anomalies, the first step proposed is to incorporate redundancy into the utilized monitoring boards.

Furthermore, a pipeline framework leveraging several algorithms is proposed to tackle these challenges. This framework detects common failures and misbehaviors, removes outliers, and defines a calibration model.

The ultimate objective of this approach is to ensure accurate and reliable monitoring in IoT networks, even when low-cost sensors are prone to single or multiple issues. This proposed pipeline provides a structured methodology to detect and rectify anomalies, thereby ensuring data integrity and reliable sensor performance. This strategy is key to maintaining and enhancing the robustness of wireless sensor networks, especially when operating low-cost sensors that might be more susceptible to malfunctions or failures. By proactively addressing potential sensor anomalies and incorporating regular calibration, the overall system performance can be optimized.

5.1 Background

The contemporary landscape of anomaly detection, as required in sensor networks, is mainly predicated on expert knowledge. A comprehensive review of anomaly detection techniques applied to IoT data is provided by Cook et al.[214], encompassing a broad spectrum of strategies while also summarizing the prevailing challenges in the domain. In a similar vein, Chen et al.[215] introduce an anomaly detection framework engineered for large, real-world sensor networks. By initially identifying spatio-temporal anomalies and regional emission sources, proceeding to rank sensing devices, and subsequently discerning malfunctioning devices, the authors affirm their framework's capability to detect outliers and infer anomalous events.

In the context of sensor calibration, various studies have contributed towards advancing our understanding and methodologies. Brattich et al.[216] characterize the performances and reproducibility of different types of low sensors in comparison to reference instruments. Moreover, the authors assess the variability of the different

sensors and perform a comparative analysis of the various optical particle counter under other meteorological conditions. Hasenfratz et al.[96, 217] undertake a unique approach by leveraging mobile sensor platforms on public transportation in Zurich, Switzerland, collecting ultrafine particle measurements over two years. This endeavor leads to the creation of pollution maps and a reduction in spatial errors. Liu et al.[39] conduct calibration on several low-cost portable sensors, demonstrating the importance of a steady particle mass concentration during the calibration process. Further, Maag et al.[65] provide a comprehensive review of state-of-the-art low-cost air pollution sensors, identifying major error sources and exploring suitable calibration models and network recalibration strategies. On a related note, Rumburg et al. [218] delve into regulatory statistics to determine the error magnitude when daily sampling is not carried out.

The aforementioned calibration processes are particularly relevant when sensor technologies that employ light-scattering techniques are considered for air quality monitoring. Budde et al.[61] juxtapose the performance of a high-accuracy measurement device with a cheap off-the-shelf sensor combined with a mobile phone, showcasing the potential of inexpensive devices through calibration and processing procedures utilizing multi-sensor data fusion. Similarly, Concas et al.[16] outline the rapidly expanding field of low-cost sensor technologies, emphasizing the role of machine learning techniques in sensor calibration. Their work also sheds light on open research challenges and future directions. In a focused case study, Crilley et al.[219] appraise the Alphasense OPC-N2, a low-cost optical particle counter, for monitoring ambient airborne particles in typical urban background sites in the UK. Their study investigates inter-unit precision, variation in measured particle mass concentration, and comparison with standard commercial optical particle counters, thus offering valuable insights into the performance and reliability of low-cost sensors.

Combining these facets—namely anomaly detection, calibration, and light-scattering techniques—it becomes evident that the future of air quality monitoring and sensor networks hinges on the integration and refinement of these elements. The enhancement of precision, reliability, and cost-effectiveness in sensor networks is achieved through the advancement of anomaly detection strategies, the improvement of calibration methodologies, and the optimization of light-scattering techniques. This lays the groundwork for a more comprehensive and accessible approach to air quality monitoring.

5.1.1 Particulate Matter Monitoring

Official Monitoring

In Italy, air pollution is regulated by the legislative decree number 155/2010, which is an actuation of the European Directive 2008/50/EC. The decree defines the minimum size and structure of the Italian territory's monitoring network, indicating the placement of stations. In addition, it specifies reference methods for measurement, analysis, and modeling of air pollutants.

Official measurement techniques for PM are based on gravimetry and β-attenuation. Gravimetric instruments use a filter to capture PM dispersed in the air sample, which is obtained via a vacuum pump. The filter is periodically replaced and weighted to determine PM mass concentrations. Size selection is performed by the air inlet, inertial impactors and the filter itself. β-attenuation devices, instead, determine the mass of PM deposit by measuring the attenuation of the radiation of a small radioactive source, when shined on the filter. Other high-precision monitoring instruments have been developed, such as TEOM and high-precision light-scattering devices. An overview PM monitoring technologies can be found in [220].

Low-cost Light-Scattering Sensors

Low-cost light-scattering PM sensors have been introduced in the market in the past a few years. They are cheaper, lighter and more compact than high-precision devices, making them suitable for IoT application. They can be sold as standalone sensors, integrated into handheld devices, or used as part of a complete IoT solution.

Low-cost light-scattering sensors work by drawing air inside the device via a small fan. A laser beam is shined on the air sample and a detector, positioned at a specific angle on the opposite side, measures the intensity of the scattered light. A deep study about this kind of sensors is presentesd in [221]. The authors classify the sensors in two different sensor technologies: nephelometers and optical particle counters. Manufactures often do not disclose which is the technology adopted by the sensors.

Nepelometers correlate the intensity of the scattered light of the whole air sample to the PM mass concentration, according to a predefined calibration curve. Optial particle counters, instead, are able to detect single particles and measure their diameter.

Particles are then classified according to their diameter in different size bins, whose number and size intervals depends on the specific implementation. By assuming spherical particles, particle density and size distribution inside each interval, the total PM mass can be computed. To obtain the PM mass concentration, this value is divided by the volume of the analyzed air sample.

Multiple assumptions and approximations are introduced by these measurement procedures. Optical properties of particles are strictly related to their Refractive Index (RI), which is not known a priori and depends on the aerosol under analysis. OPCs use Mie Theory, that models light scattered by a perfect sphere, in order to measure particle diameters. However, the theory assumes spherical particles, which is not true for real-world PM, and a known RI. In addition, the conversion from PM volume to mass also requires the knowledge of PM density.

Therefore, calibration curves of both nephelometers and OPCs strictly depend on the particulate type that is used during the calibration procedure. Artificial PM composed of Polystyrene Spherical Latex Particles (PSLs) of a known diameter and RI using are often used. Alternatively, more realistic but less comprehensive PM compositions, such as cigarette smoke, can be utilized.

Low-cost devices may not detect the full number of particles in the sampling volume relying on statistics and extrapolation to compute their actual number. Accuracy becomes worse for increased particle sizes since their number decreases dramatically. For this reason, bigger particle concentrations, such as PM10 and PM4, are often estimated from PM1 and PM2.5. Particle detection is also limited by a mininum size threshold.

High levels of relative humidity greatly affect low-cost particulate matter sensors. Hygroscopic growth both increases the size of the particles, leading to an overestimation of PM mass, and influences their optical properties. Full-size particle counters solve this problem by heating the air before performing the measurement.

5.1.2 Experiment Deployment Area

Turin is situated in the North-West of Italy and is the capital city of the Piedmont Region. This area is surrounded by the Alps on the West and North side and by a big hill on the East side (Figure 5.1). Each location of the ARPA air quality monitoring stations are classified according to the level of urbanization, the level of urbanization

include urban, suburban, and rural areas [108]. The local environmental protection agency provides annual reports on the air quality in the city.

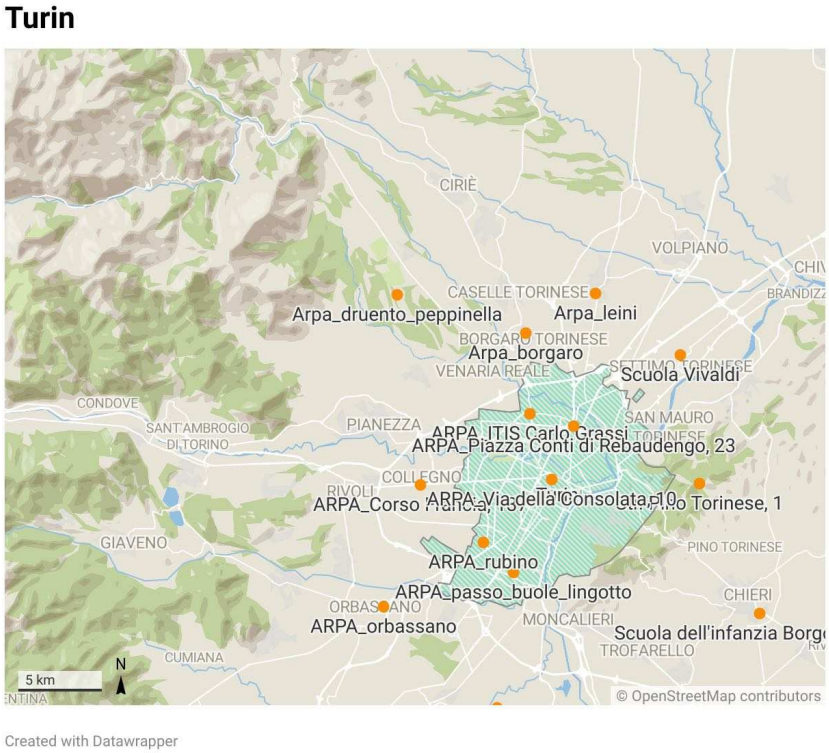


Fig. 5.1 Metropolitan area of Turin.

The official measurements of PM2.5 at the Torino-Rubino station at the time of the experimental campaign are shown in Figure 5.2. The trend shows an increase in background PM concentrations in winter, starting mid-October. From late winter/early spring, PM levels decrease until they reach a minimum during the summer period. The trend similarly repeats itself every year.

The increase in PM2.5 in the winter period, with respect to the rest of the year, can be attributed to heating systems. This situation is worsened by thermal inversion, a phenomenon mainly present in winter, that causes air stagnation. Thermal inversion happens when there is a temperature stratification of the air above the city, where temperature increases with altitude. This phenomenon prevents air circulation and dispersion of pollutants [222].

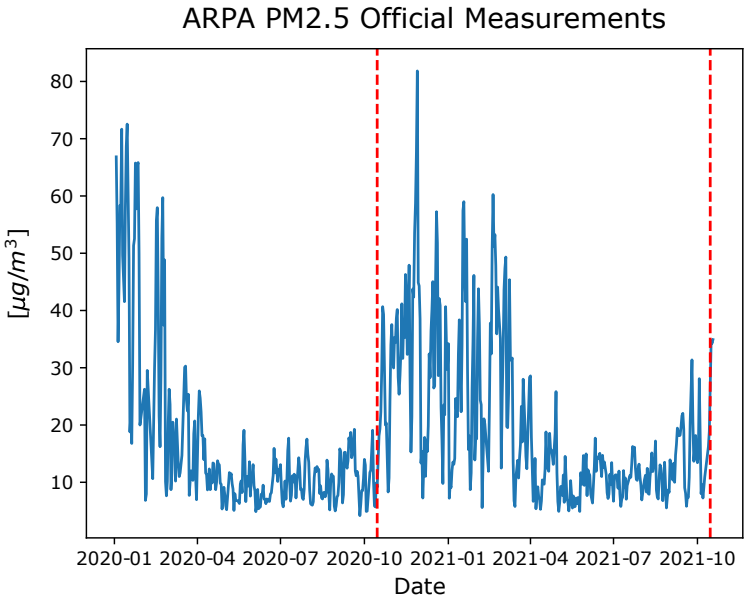


Fig. 5.2 Official data of the yearly trend of $PM_2.5$ density at Torino-Rubino. In red is labeled the start and end period of the experiment period presented in this chapter.

5.1.3 Experimental Campaign

The study presented in this chapter uses the air monitoring stations presented in Chapter 2.2.3, these stations employ a Raspberry Pi Zero Wireless as the single-board computer running essential components of Arch Linux operating system for ARM. The system collects data from various sensors, including the Honeywell HPMA115S0-XXX sensor for $PM_{2.5}$ and PM_{10} particles, utilizing the Light Scattering technique for particle concentration estimation. The DHT22 sensor provides temperature and humidity readings, and the BME280 sensor from Bosch captures atmospheric pressure measurements. Python scripts query the sensors, and specific APIs enable sensor-board communication through various protocols.

In this context, fourteen monitoring stations were situated on the rooftop of the Rubino ARPA station, as illustrated in Figure 5.3. These stationary sensors were maintained at the ARPA reference station location for a period of one year. The sensors were positioned 1.5 $2m$ away from the air inlet of the reference grade device on the premises. The monitoring station was strategically placed in a public park, partially distanced from vehicular traffic.

Fig. 5.3 Setup of Air Monitoring Stations on the Rooftop of the Rubino Station

5.2 Data Improvement Framework

Fault detection and sensor calibration are critical tasks that can affect the reliability and performance of PM measurement [80]. As described in the previous section, sensors are based on several working principle that can generate measure outliers, due to environmental conditions or manufacturing variations. During our measurement campaign, we observed that some sensors exhibited persistent outliers, suggesting a potential failure in the sensor. These failures significantly affect the calibration process, which relies on accurate and consistent measurements to generate a precise fitting model. When the sensors are not calibrated correctly, the measurements they generate can be erroneous, leading to data noise or inaccurate readings that can impact the overall performance and reliability of the system. To mitigate these issues, we propose a novel framework that leverages algorithms to detect the most common failure in the sensors, removing outliers and calibrating them automatically, providing more accurate measurements even in the presence of outliers or failures.

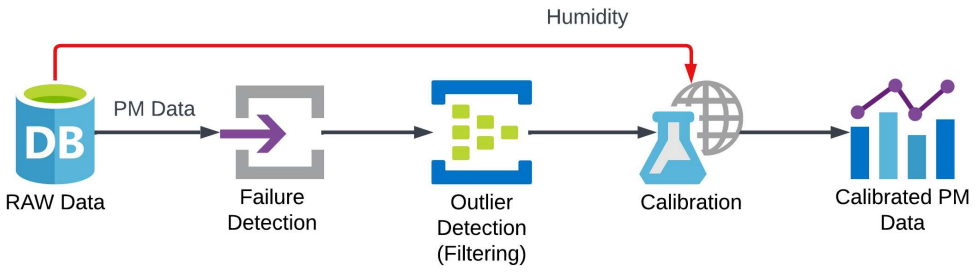


Fig. 5.4 Framework Overview: Data improvement data flow

The air pollution monitoring stations, throughout the course of the measurement campaign, deposit raw PM and humidity values from various sensors. The method outlined in this paper commences with the retrieval of raw PM and humidity measurements that have been collected by these air pollution monitoring stations. The raw data of PM is processed using a fault detection algorithm, followed by applying a filtering technique for the removal of data outliers. With these preprocessed data, ground truth data and the humidity measures are taken for the PM sensors calibration.

Upon finalization of the process, the calibrated model is applied taking into account the associated humidity values. The median PM value is subsequently computed from the sensors within each air pollution monitoring station, aiming

to establish the particular PM measurement for each station. To present a more comprehensive exposition of our methodology, the ensuing sections delineate each algorithm engaged in the calibration and deployment processes, highlighting their contribution to the assurance of precise and reliable PM measurements.

5.2.1 Failure Detection

During this step, the raw readings obtained from the sensors are analyzed to identify any sensor failure. A failure can occur for various reasons, such as a malfunctioning sensor or a faulty connection. Faults in the embedded photosensor are the primary cause of malfunctioning in these sensors, a fact that remains obscure due to the sensor's uninterrupted data production, absent of any warning signals. This lack of obvious failure signs makes detecting the fault a challenging task. Hence, it necessitates a thorough inspection of the generated measurements to ensure effective fault detection.

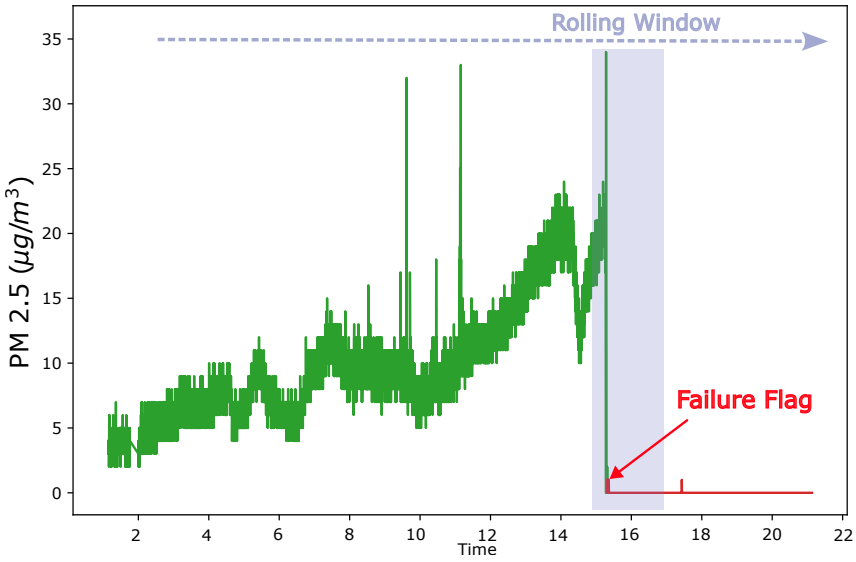


Fig. 5.5 Fault detection process. Particulate matter density($\mu g/m^3$ vs Time (hours) plot. The algorithm utilizes a sliding window approach through the data to identify periods with anomalous values, subsequently marking the failure point.

The algorithm implements a rolling window technique, as depicted in Figure 5.5, to identify when the values remained at their lowest values (0 or 1 $\mu g/m^3$). If the sensor's data remained at these values for the entire length of the window, the

sensor was flagged as damaged. Any data obtained after the flag was disregarded for subsequent steps.

To optimize this process, multiple time windows were tested, ranging from 1 to 12 to hours. The aim was to determine the ideal window that would offer the highest accuracy in detecting actual events while minimizing false positives. This step was crucial in ensuring the reliability and accuracy of the subsequent analyses.

5.2.2 Outlier Detection

Once the data has been analyzed for failures, the next step in our pipeline requires filtering out the outliers. In this phase, the goal is to remove all readings that are affected by noise or external factors. Please note that outliers can significantly skew the data and lead to inaccurate measurements and it is essential to remove them before calibration.

Those events produce high-frequency changes, often impulsive, which tend to increase the errors and lead to data outliers. They can also lead to rapid and sustained changes in PM values for several minutes. As a consequence, our filter must be designed to effectively reduce the number of outliers and minimize errors with respect to the ground truth values. We perform several tests to identify the optimal filter removing the noise and maintaining the accuracy. During these tests, we take into consideration several factors. First, some errors may have a random nature, resulting in high-frequency noise, whereas others may be impulse noise. Thus, a low-pass filter was deemed appropriate to eliminate this kind of errors. In contrast, we select a non-linear filter (a median filter) to analyze the effect of a particle stuck inside the sensor. To ensure that the signal remained unchanged in the band-pass and to avoid a stringent cutoff band that would alter the signal shape, we select an eighth-order Butterworth IIR filter and we also implemented as a low-pass filter. This filter was chosen to achieve a slower roll-off, which helps to preserve the signal's shape. The filter was implemented using the Python `scypi` signal processing library, which provided the necessary functionality. Similarly, for the median filter, we use the `scypi` library for its ability to implement the filter easily and accurately. The cut-off frequencies and kernel sizes for both filters were selected according to the values described by Espinosa et al. [56].

Based on an initial manual examination, it has been observed that certain erroneous measurements occur within time intervals of a few minutes. Consequently, these measurements cannot be removed by the filters using the predefined setting values established previously.

Despite the effectiveness of these filters, it was acknowledged that their determination of an outlier was not entirely effective. To address this limitation, we employ the Z-score filter to remove signal outliers.

The Z-score filter is a statistical filter used to remove outliers in a data set based on the standard deviation and mean of the set. To preserve the overall shape and structure of the signal, we flag data points outside a certain range of the standard deviation. The Z-score is calculated using the formula:

$$z = \frac{x-\mu}{\sigma}$$

where μ represents the mean value and σ the standard deviation of the data set. The resulting Z-score z represents a scalar number of how many standard deviations a data point is away from the mean. Positive values indicate points above and negative values points below the mean.

Based on a threshold value, every point with an absolute value of the Z-score above the threshold is considered an outlier and is either removed or replaced. However, for PM data, which may exhibit significant variations over time, defining static mean and standard deviation values is not appropriate. Therefore, we propose a time-dependent windowing of the database to avoid false positives in the Z-score filter. Specifically, z is computed based on the time period around the data point being evaluated. This approach can accurately remove outliers and preserving data integrity. To sum up, during the filtering phase it was critical to ensure that the sensor readings were reliable and reflected pollution conditions. Thus, to select and apply the most appropriate filter we consider the unique characteristics of the noise and the outlier patterns.

5.2.3 Calibration Model

The final step in the deployment process is sensor calibration. During this stage, the readings coming from the previous steps are used to calibrate the sensors, using

the model obtained during the calibration phase. The calibration algorithm uses the raw data coming from the sensors, i.e., PM and humidity data. Once the calibration model is applied, the median of the calibration outputs of all sensors is taken as the valid PM value, and the performance of the complete framework is evaluated.

The calibration model presented in [13] is used as a reference, in which it is shown that for this type of sensor and platform, calibration through multivariable linear regression displayed the best performance. In addition to this model, data calibration is performed using both filtered and unfiltered data as discussed in section 5.2.2. Moreover, a new scheme is introduced, in which other anomalous values detected by a modified Gaussian mixture algorithm are omitted to reduce their influence during the calibration period. This model is added to the reference calibration using both filtered and unfiltered data and the process of outlier detection. These four models are then analyzed to determine the best performing model.

The final step in the deployment process is sensor calibration. This work tries develop a calibration process that doesn't require human intervention. This is due to the fact that in large scale deployment scenario the amount of data produced by the sensors would be too high to be manually analyzed. As shown in [13] for this type of sensor and platform, multivariate linear regression, using PM2.5 and relative humidity as independent variables, provides good performance.

The calibration process is logically divided in two time phases. In the first phase, sensors are positioned at a reference stations and data coming from both the sensors and the reference station is collected. The data acquired during this phase is used to train the calibration model. In the second phase, the calibrated model can be used to correct the sensors in real time.

The measurements collected during the first three weeks of the experiment, from October 10 to October 31, 2020, were used for training the model. Then, the trained model was used to calibrate sensor data acquired during the remaining part of the experiment campaign, to simulate a real deployment scenario. The performance of the calibrated sensors is evaluated with respect to the reference instrument. Since the finest data granularity available from the reference station are hour measurements, the calibration process is performed on hour aggregations of the sensors' readings.

In order to enable automatic calibration of the sensors, it is essential that the model is correctly trained. It was observed that the model over-compensated the correction, by selecting excessively small PM2.5 coefficients. Liner regression

models are known to be heavily influenced by outliers, so this behavior could be partially attributed to the presence of sporadic measurements that are much higher that the reference. In addition, the removal of outliers can provide a calibration model that more general, without being influenced by occasional sensor faults occurring during the calibration period.

For this reasons, an automatic outlier-detection method, based on multivariate gaussian model (GM), is introduced. The objective is to select the data point to remove from the dataset used to train the calibration model. The proposed method consists in fitting a multivariate gaussian distribution to the 2-dimensional data points composed of the PM2.5 measurement of the sensor and the corresponding value from the reference station, the latter being available when training the model.

The next step is to set a threshold probability. A cumulative probability function for the 2-dimensional gaussian distribution can be defined in function of the Mahalanobis distance from the sample mean:

$$dist(x_i) = \sqrt{(x_i - x_{mean})V^{-1}(x_i - x_{man})^T}$$

Intuitively, the farther the point are from the sample mean, the lower is the probability of being measured. Given a data point, if the probability of measuring values at a distance greater or equal of the one of data point is lower than the threshold, the data point should be removed. This probability can be computed as follows [223]:

$$p(x|dist(x) \geq dist(x_i)) = e^{-(dist(x_i)^2)/2}$$

Using a probability threshold instead of removing a fixed percentage of the less probable data, allows the models to better adapt to changes in behavior between the sensors.

However, according to the literature [218, 217] and to what was observed for the collected data, measurements coming from both the sensors and the reference station are better represented by a lognormal distribution, rather than a gaussian. For this reason, before applying the outlier detection model, the hour measurements of the reference station and each single sensors are fitted with a lognormal distribution using the `scypi` software package (`scipy.stats.lognorm.fit`).

The fitting allows us to estimate the shift parameter of the lognormal distribution, and to apply a transformation to the data so that it follows a normal distribution:

$$normal = ln(lognormal - shift)$$

Once the data is transformed, the multivariate gaussian model is applied as previously described to remove outliers. Finally, the remaining data is transformed back and used to train the calibration model.

For the analysis presented in this work, the threshold probability was set to 5%. In addition, sensors showing a correlation with the reference lower than 0.65 were also removed. The reason is that if the sensor is not functioning properly during the calibration period, it should be identified and discarded.

5.2.4 Experiment Methodology

To evaluate the performance of our pipeline framework, all sensors data were manually inspected to determine sensor functionality after conducting the experiment described in Section 5.1.3. Sensors were classified into i) those that functioned properly, ii) those that failed from the beginning, and iii) those that failed during the measurement period. This classification was then used to evaluate the failure detection performance of the framework for both processes.

Once we classified the sensors, we compare the readings of each board with the ground truth reference coming from ARPA. However, ARPA provides new measures every hour, whereas the our sensor network provides new measures every second. To obtain comparable measures, we aggregate per-second data into the equivalent hourly average. The resulting values were then compared with ARPA reference values to analyze the precision of our system. The comparison was performed for both the raw input data and the framework output data to evaluate the level of precision enhancement provided by the proposed framework.

To assess the contribution of each process within the framework, each step is evaluated under different parameters with the goal of identifying the optimal set. Subsequently, a collective evaluation is conducted to determine the level of data accuracy improvement in comparison to reference values.

Table 5.1 Summary of framework's performance

Metric	Framework Step			
	Failure Detection	Outlier Detection	Calibration Model	Global Performance
RMSE		✓	✓	✓
MAE		✓	✓	✓
r^2		✓	✓	✓
R^2		✓	✓	✓
Confusion Matrix	✓			

For the fault detection process, a confusion matrix is analyzed to identify the window type that offers the highest precision in detecting sensor faults. In the anomaly detection and calibration processes, performance is assessed through the analysis of error metrics (RMSE, MAE) and correlation in relation to the reference values (r^2 and R^2). Table 5.1 provides a summary of the performance metrics that the framework evaluates.

5.3 Framework Performance

After the measurement campaign, the 56 low-cost PM sensors employed in the 14 air monitoring boards were manually inspected to determine their functionality. We identified the sensors that worked properly (without significant issues of long-term erroneous data, highly noisy measures, or values fixed in the lower range of the sensor scale) and the ones that stopped working due to random and uncorrelated measures.

The inspection showed that 13 PM sensors (23.21%) functioned properly throughout the experiment's duration, whereas 10 sensors (17.86%) failed at the beginning of the experiment, and 33 sensors (58.93%) started to fail during the experiment. As we did not notice any electronic issues during data logging or transmission to the main processor, malfunctions were typically due to errors in the optical part of the sensor. This statistic confirms the lack of reliability of low-cost PM sensors when exposed to long periods of outdoor operation. It also shows the importance of identifying spurious PM values and highlights the necessity to adopt a framework capable of

automatically detecting and removing problematic sensors and their readings, such as the one proposed in this study.

In the evaluation process, it was determined through manual comparison that the reference instrument employed by ARPA does not register values below 4 $\mu g/m^3$. Therefore, when evaluating the framework's performance, the inclusion of values equal or below leads to an increase in error and negatively affects performance measures. Given the minimal magnitude of these measurements, values equal or less than 4 $\mu g/m^3$ are excluded from the evaluations.

In this section, we follow the description flow used in Section 5.2, i.e., each stage of our pipeline framework is evaluated to determine its contribution to the final performance.

5.3.1 Failure Detection

The random and noisy behavior exhibited by the PM sensors poses a challenge when defining a time window that accurately detects sensor failures. A window with an overly short period might generate a high number of false detections, especially during the summer season when PM levels are typically lower. On the other hand, very long windows can omit erroneous behavior of the sensors, as explained in Section 5.1.1, leading to a loss of sensitivity and generating false negatives. We experimented with six different time windows, including 1, 2, 4, 8, and 12 hours, based on empirical observations. In order to ascertain false positives and false negatives in the detection process, we define an erroneous detection as the one in which the algorithm identifies a failure after two-weeks with respect to the manual inspection.

The fault detection results for each time window of the 56 sensors are presented in Table 5.2. Table 5.3 shows the metrics used to evaluate the algorithm's performance. These results emphasize the inherent trade-off between accuracy and sensitivity. Indeed, shorter windows may be prone to false positives when the particulate matter level is low, whereas longer time windows reduces the probability of detection. Notably, the windows lasting 8 and 12 hours demonstrate superior performance in the fault detection task. Moreover, adopting time window longer than 12 hours leads to an increase in false negative detections. Table 5.3 also shows that the 12-hour window, albeit less sensitive, yields fewer errors in fault detection, as indicated by the

Table 5.2 Summary of Confusion Matrices

		True Positives	**False Positive**	**False Negatives**	**True Negatives**
Window Size (hours)	1	21	23	7	5
	2	28	13	7	8
	3	31	7	9	9
	4	32	5	9	10
	8	32	3	10	11
	12	32	0	11	13

Table 5.3 Confusion Matrices Metrics Comparison

		Accuracy	Precision	Recall	F1-Score
Window Size (hours)	1	0,464	0,477	0,750	0,583
	2	0,643	0,683	0,800	0,737
	3	0,714	0,816	0,438	0,570
	4	0,750	0,865	0,780	0,821
	8	0,768	0,914	0,762	0,831
	12	0,804	1,000	0,744	0,853

higher F1-Score value. The faults that were not detected within this window pertain to sensors that exhibited highly noisy data before stabilizing at their lowest values or sensors that consistently measured noisy values throughout the entire measurement period. Detecting such faults is challenging and necessitates additional analysis to accurately determine the underlying cause.

Figure 5.6 deepens our analysis by plotting the percentage of invalid readings with respect to the total number of values. Positive values on the y-axis denote the percentage of valid information that is removed due to false positives, leading to the exclusion of genuine PM data. On the contrary, false negatives result in the incorporation of erroneous information into the framework, thereby influencing its performance. An optimal detection scenario is indicated by a value equal to zero. To assess the relative effectiveness of the different time windows, we also compute the root mean square error of all values. Notably, the 12-hour window once again demonstrates the closest approximation to the ideal scenario.

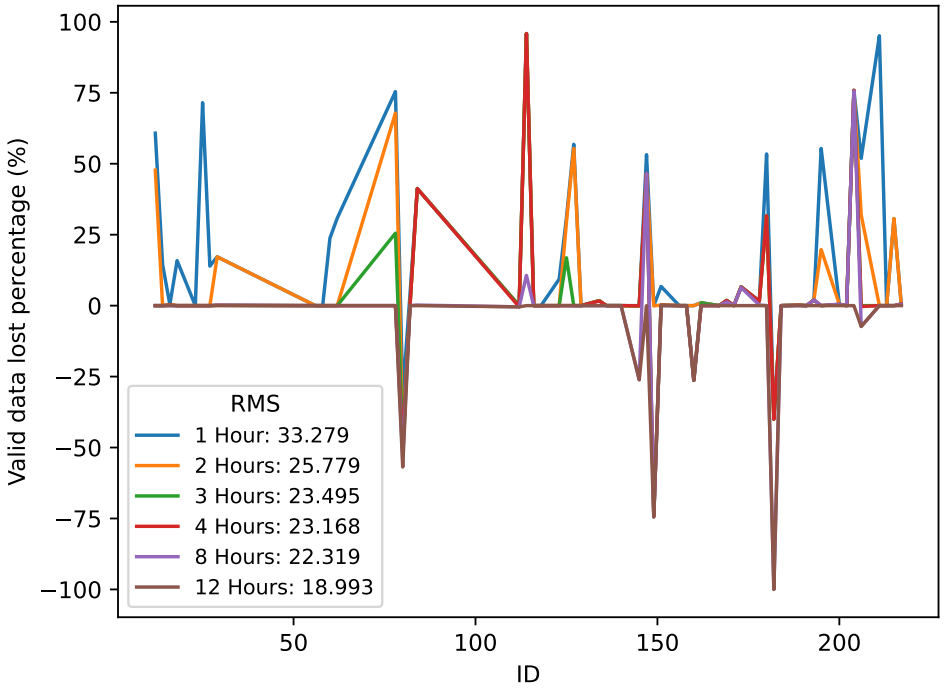


Fig. 5.6 Valid data loss for each sensor ID in different time windows. Negative values indicate erroneous data that was not removed after a failure. A root mean square (RMS) value close to zero represents the ideal scenario.

5.3.2 Outlier Detection

Following the window detection phase, various filters are assessed for the detection and elimination of anomalous values that impair the measurement. At this stage, the reduction in error and its impact on the correlation index compared to reference values are analyzed after the application of the proposed three filters. Additionally, during the data aggregation process, a comparison is made of the percentage of summarized hourly data points that show a significant change in value, either decreasing error (*% Points Corrected*) or increasing it (*% Points Deprecated*). Table 5.4 summarizes the evaluation of these parameters for each of the filters, each value representing the global mean value of measurements for each sensor over the experimental period.

Based on these values, the Z-score filter demonstrates better performance in terms of error reduction and an improvement in the Pearson's coefficient (r^2) value, as well as a greater impact on the percentage of summarized data points that are

corrected. On the other hand, the median filter shows the second-best performance, prompting an evaluation of the combination of these two filters (Median + Z-Score) applied in a cascade manner. However, these processes exhibit worse performance than each of the other filters when applied individually. Given these findings, the Z-score filter is selected for inclusion in the framework pipeline to detect and remove anomalous values prior to the application of the calibration process.

Table 5.4 Outlier Detection: Filters performance comparison

	RMSE $\mu g/m^3$	MAE $\mu g/m^3$	r^2	% Points Corrected	% Points Deprecated
Raw Data	19,081	13,723	0,792	-	-
Low Pass Filter	19,122	13,734	0,790	0,992	0,531
Median Filter	18,975	13,636	0,793	0,441	0,160
Z-Score	18,843	13,607	0,805	2,511	1,831
Median + Z-Score	18,439	14,711	0,781	1,815	2,040

5.3.3 Calibration Model

In order to conduct an in-depth evaluation of each calibration model, the sensors that exhibited failures but were not identified by the algorithm were intentionally excluded from the calibration performance analysis. This decision was made to streamline the metrics and provide a more accurate depiction of the overall performance, eliminating the interference that these failed sensors might have introduced.

Figures 5.7 and 5.8 display the respective distribution of error metrics (RMSE), and the correlation coefficient (R^2). Upon the detailed examination of the results, it was observed that the calibration models incorporating a filtering process for outlier detection, exhibited superior performance metrics as compared to the base reference model (PM + Hum) presented in [13]. This suggests the efficacy of outlier detection in enhancing model calibration.

Additionally, the use of the calibration model with the outlier-detection based on multivariate Gaussian model minimizes the impact of anomalous or highly variable values (see PM + Hum + GM, Filtered PM + Hum + GM). The implementation of

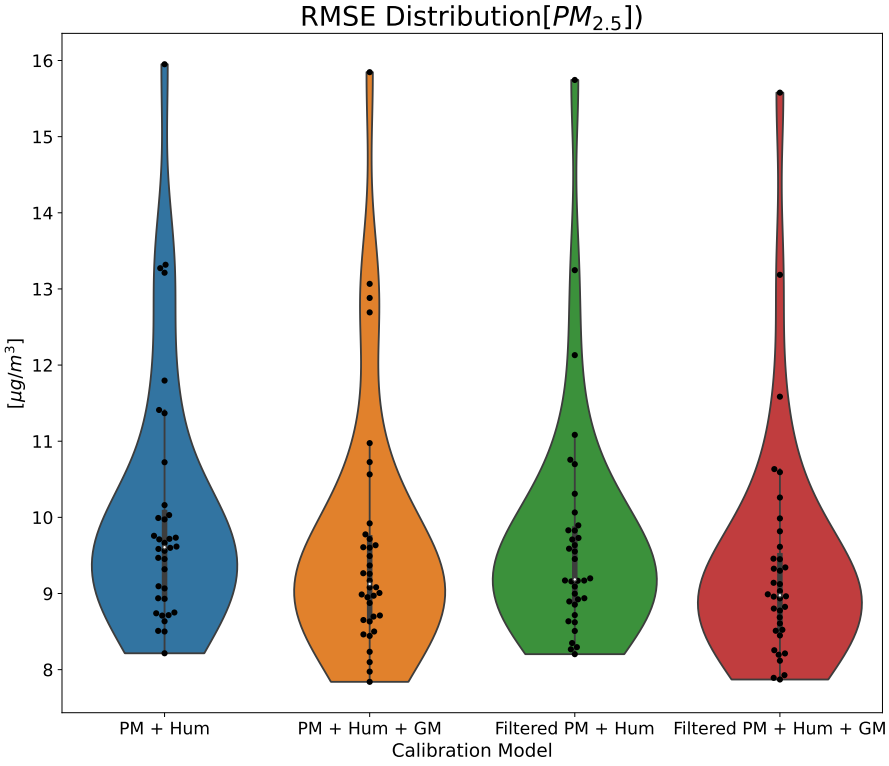


Fig. 5.7 RMSE Error distribution for each calibration model, the points represents the average error of each sensor.

this algorithm, in tandem with the other methods, led to a calibration model that proved to be more robust and resilient to errors.

Table 5.5 encapsulates the performance metrics for each calibration model, providing a global view of their performance. This table showcases the median values registered by the sensors throughout the experimental period. A significant observation was that the implemented filtering process had a substantial role in improving the correlation between variables, while the GM algorithm was successful in decreasing the error value.

Ultimately, the filtered PM + Hum + GM model demonstrated the highest efficacy in calibration, with notable improvements in both error reduction (RMSE and MAE by 6.6%) and an increase in the correlation index (R^2) by 4.5%. The amalgamation

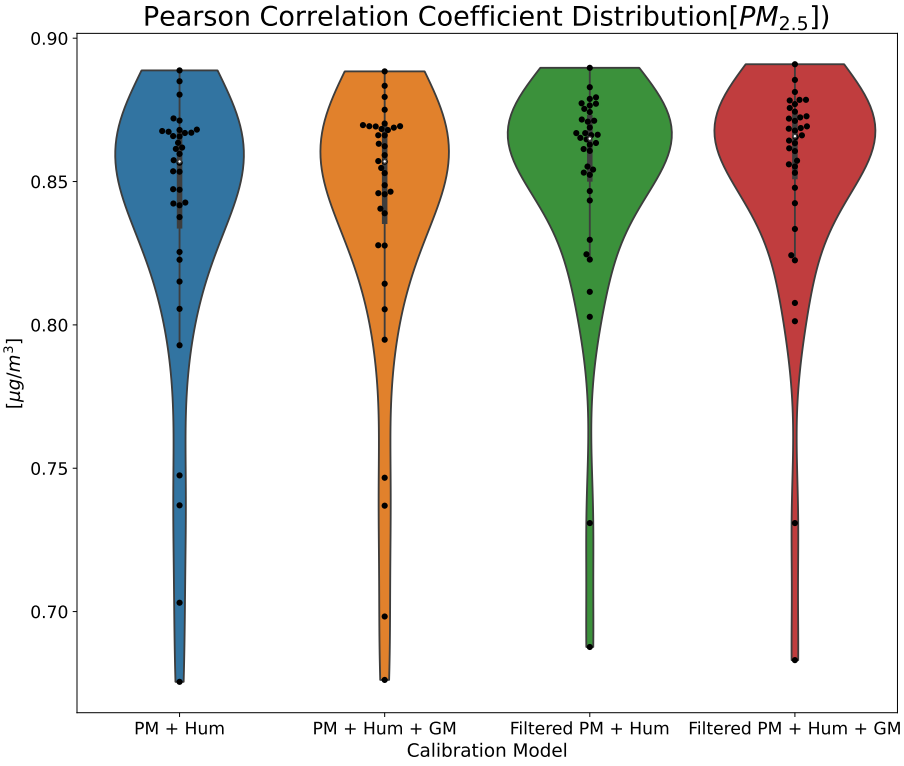


Fig. 5.8 Pearson's correlation index (r^2) distribution for each calibration model, the points represents the average r^2 of each sensor.

of these improvements underscores the value of the combined use of outliers filter and GM algorithm in optimizing calibration methods.

5.4 Global Performance

In our comprehensive evaluation of the proposed measurement framework, we've considered each component of the process chain in unison, encompassing the sensors integrated within each of the 14 distinct monitoring stations. For each of these stations, the determination of the hourly median provides an aggregate measure of performance across all operational sensors within the given period. This approach allows us to capture the most reliable and representative data from an ensemble of PM sensors, thus boosting the robustness of our analysis. These resulting measures

Table 5.5 Calibration model performance comparison metrics

	RMSE $\mu g/m^3$	MAE $\mu g/m^3$	r^2	R^2
PM + Hum	9,608	7,149	0,857	0,670
PM + Hum + GM	9,125	6,744	0,857	0,692
Filtered PM + Hum	9,185	6,960	0,865	0,677
Filtered PM + Hum + GM	8,976	6,681	0,866	0,700

then serve as the foundation upon which performance metrics are calculated and compared against the reference calibration model. We've summarized these all-encompassing metrics in Table 5.6, providing a holistic overview of our system's performance.

Our findings clearly illuminate the efficacy of our system in controlling the impact of false negatives. Thanks to the inbuilt redundancy of sensors within each monitoring station, the majority of cases saw the system demonstrate either equivalent or superior performance when compared with the values presented in Table 5.5. This observation is a clear indicator of the robustness of our framework, showcasing its ability to enhance the precision of measurements despite the challenges posed by the high failure rate of these sensors observed over the course of the year-long deployment period.

Table 5.6 Global framework performance metrics vs reference model

	RMSE $\mu g/m^3$	MAE $\mu g/m^3$	r^2	R^2
Reference Calibration Model PM + Hum [13]	9,006	6,757	0,858	0,654
Filtered PM + Hum + GM	8,462	6,241	0,868	0,725

Analyzing the distributions portrayed by the box plots in Figures 5.9 and 5.10, we observe that the dispersion in the measurements remains remarkably stable. This observation provides evidence of high coherence and consistency in the measurement process, further bolstering the reliability of our framework. Also the median is under the error values provided by the sensor's data-sheet (RMSE aprox of 18.4) [102].

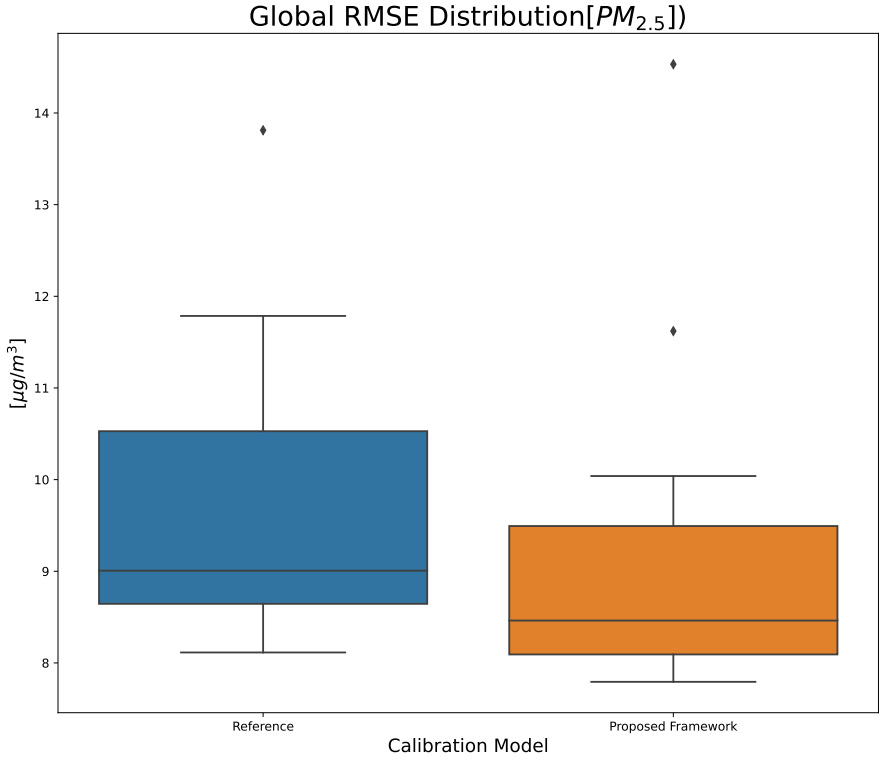


Fig. 5.9 Comparison of RMSE distribution for all monitoring stations between the proposed framework and the reference calibration model.

However, the framework is not without its limitations. Notably, the presence of more than two faulty sensors within a single station can pose challenges to the system's reliability. This occurrence, while relatively rare, can result in outlier points in terms of erro and the correlation coefficient ($R2$) as demonstrated in Fig 5.10). These anomalous points, although infrequent, underscore the importance of sensor redundancy and robust error handling mechanisms in ensuring the overall performance and reliability of the system.

5.5 Conclusions

Anomaly detection involves the identification of malfunctioning and failures which significantly affect the overall performance and reliability of a sensor network. In

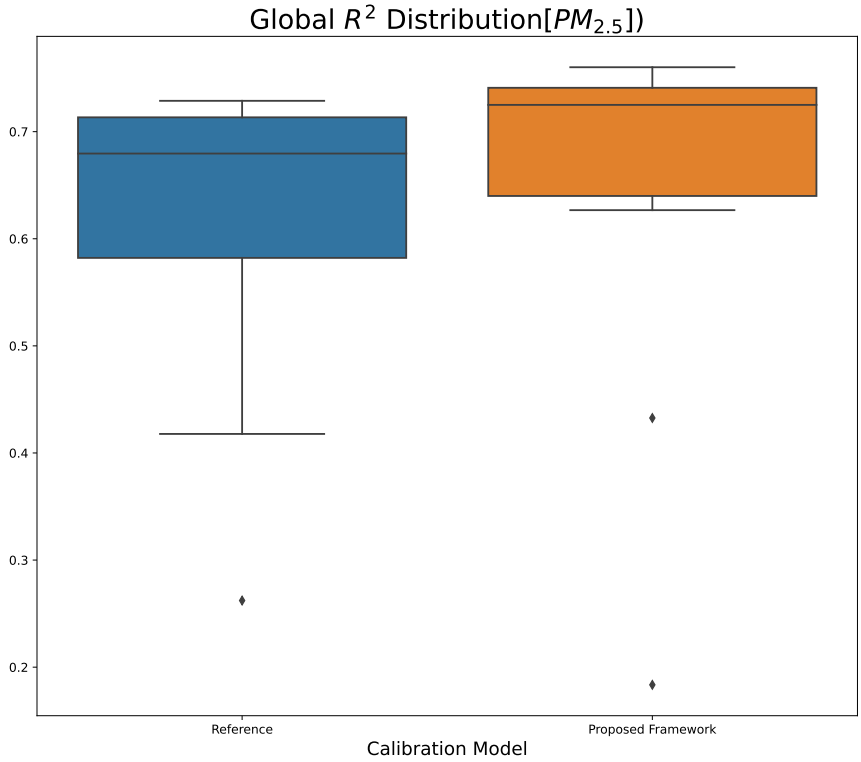


Fig. 5.10 Comparison of the coefficient of determination (R^2) distribution for all monitoring stations between the proposed framework and the reference calibration model.

this work, we propose a pipeline in which each stage is dedicated to detect some type of anomaly. The overall framework leverages several algorithms to detect the most common failure in the sensors, remove outliers, and calibrate the sensors.

The system delivers more precise measurements, even in the face of anomalies caused by partial or total sensor failure. Furthermore, it reduces dispersion in the measurements, thereby increasing coherence in these types of sensors compared to other reference calibration systems. It also enables the measurement median to fall below the error levels specified by the manufacturer.

On a broader scale, the obtained results present a promising outlook for these types of sensors in the field of particulate matter measurement within a smart city environment. Fault detection also paves the way for maintenance or replacements that ensure the sensor's longevity for extended measurement periods. In environments

like cities where climatic conditions are more stable (near the Earth's equator), this feature opens up the potential for continuous monitoring. This contrasts with cities like those discussed in this thesis, which experience critical conditions at specific times of the year.

Chapter 6

Data Integrity

The recent ascendancy of blockchain and other Distributed Ledger Technologies (DLTs) has permeated not only the realm of cryptocurrencies but also myriad application areas that demand utmost data integrity and authenticity. The evolving landscape of smart cities embodies these necessities, given that transparency in service delivery to citizens increasingly holds the key to fostering public trust.

The present chapter commences with the introduction of a proposed architecture for a biometric electronic identification document (e-ID) system, underpinned by Blockchain, for citizen identity verification during transactions related to notary services, registration, tax declarations and payments, primary health services, and economic activity registrations, among others. A biometric e-ID system is leveraged to confirm user authentication, thereby mitigating the risks of spoofing and associated attacks. Moreover, to validate the document, a digital certificate, supplemented with corresponding public and private keys for each citizen, is employed, accessible via a user's PIN.

This proposed transaction validation process was actualized on a Blockchain system to record and verify transactions executed by all citizens registered in the electoral census, thereby ensuring security, integrity, scalability, traceability, and non-ambiguity. Furthermore, a Blockchain network architecture is exhibited in a distributed and decentralized fashion, encompassing all network nodes, databases, and governmental entities such as national registers and notary offices. The outcomes of applying a novel consensus algorithm to our Blockchain network are also disclosed,

highlighting the mining time, memory, and CPU usage as the transaction volume escalates.

Following this, the chapter pivots to the presentation of a proof of concept, accompanying a case study on particulate matter pollution measurement, based on the benefits and results proffered by blockchain. This segment discloses an architecture that facilitates the integration of an Ethereum-based blockchain with the architecture delineated in Chapter 4. The proposed solution adeptly fuses blockchain storage through a marriage of centralized storage in a database and distributed storage via blockchain. A noteworthy feature introduced is the intertwining of hash codes, enhancing data immutability and security, and enabling users to conduct verifications, thereby reinforcing their confidence in the open data made available by the architecture in terms of particulate matter density.

Some of the work described in this chapter has been previously published in [224]

6.1 Blockchain and Smartcities

The increasing integration of ICT into our everyday lives has brought about a new paradigm for modern societies. This technological advancement has also sparked a heightened interest in the application of technology in governmental processes, leading governments to prioritize the implementation of a novel concept known as electronic Government (e-Government). Various studies have defined e-Government as a strategic approach that utilizes ICT applications to enhance public services and facilitate greater interaction between the government and its citizens, employees, or internal entities. The overarching goal of e-Government is to improve efficiency and effectiveness within the government itself, as well as in its interactions with other governmental entities [225].

The definition of e-Government encompasses four primary dimensions: i) Government to Citizen (G2C), ii) Government to Company (G2B), iii) Government to Employee (G2E), and iv) Government to Government (G2G) interactions [225]. One notable aspect of e-Government is its utilization of ICT for citizen registration and identification purposes. It is important to clarify that in this study, the concept of identification refers to the verification of an individual or entity's identity. In

Latin American countries, for instance, a registration process is required, which simultaneously represents the constitutional right of citizens to be uniquely identified by the state from birth [226]. The official registration process culminates with the issuance of an Identity Document (ID), which serves as a distinct identifier in the government registry database [227]. Typically, the ID contains personal information such as the bearer's photograph, name, date of birth, and other relevant details. This identification process acts as a means to verify a person's identity. However, with the advancements in technology, a novel concept of Electronic Identity Document (e-ID) has emerged. The e-ID involves generating the same ID in a smart card format, where the carrier's data can be digitally stored, including additional features such as facial and fingerprint information for recognition. Furthermore, it incorporates enhanced security measures through the encryption of personal information, granting citizens access to online services [228].

Furthermore, the utilization of blockchain technology (BCT) presents a significant opportunity to address various public challenges, particularly those associated with corruption. Implementing BCT in the government sector can have a profound impact, especially in scenarios where citizen demand transparency from their government, such as in public data management, electronic voting, and taxation processes. Currently, the national identification (ID) systems used in many countries face security issues, with authorities reporting thousands of lost documents, which can potentially lead to identity theft. There have also been instances where criminals have cloned individuals' documents to evade authorities and engage in fraudulent activities. Additionally, cases of lost or stolen identification have been exploited by criminals to falsify people's identities, enabling them to obtain bank loans or fraudulently sell properties [229].

Given the existing vulnerabilities and security risks associated with national ID systems, particularly in transactional contexts, there is a clear oportunity for the implementation of ICT technologies to monitor and authenticate these transactions. The utilization of blockchain networks, in conjunction with robust authentication and user identification processes, offers a potential solution to address the aforementioned issues. In the context of ID systems, user authentication refers to the process of verifying and validating the connection between an ID document and its rightful owner. This process is based on the concept of strong authentication, as proposed in [230], which encompasses three factors: possession of something (e.g., a card), knowledge of something (e.g., a PIN), and being or doing something (e.g., fingerprint

or signature). The combination of these factors results in a more secure and robust authentication method for verifying an individual's identity.

The objective of this research is to put forward an architectural framework for transaction management that incorporates biometric authentication and a national electronic Identity Document (e-ID) system. The proposed architecture leverages Blockchain technology for transaction validation, utilizes smart cards, and integrates biometric user authentication. It is hypothesized that the functionalities of BCT can be validated in common government transactions, and security measures for user authentication can be improved by including fingerprint and iris recognition information into a smart card.

6.1.1 Related Work

BCT is grounded on the concept of a decentralized ledger that cannot be modified or tampered with, requiring consensus from all network members and recording all validated transactions. This technology offers decentralization, integrity, reliability, information traceability, and non-repudiation by users. These functions bring various benefits, including transaction veracity [231]. In a Blockchain, all participating nodes have access to the registered information at any given time. Therefore, if a node is under attack, other network members can identify it and reject the counterfeit block, thereby preventing fraudulent transactions [232, 233].

Despite being hailed as a promising technology for secure transactions, BCT has undergone scrutiny, revealing several security issues that have been thoroughly examined [234, 235].

- *Balance Attack*: Specifically targeting BCT systems that use the proof of work (PoW) consensus algorithm, this attack enables double-spending [236].
- *Selfish Mining Attack*: This attack aims to gain undeserved rewards or waste the resources of honest miners [237, 238].
- *Eclipse Attack*: This attack involves gaining control over all the connections of a victim, isolating them from other peers and serving as a foundation for further attacks [239, 240].

- *Decentralized Autonomous Organization (DAO) Attack*: Malicious smart contracts are deployed on a smart contract platform, exploiting its functionality for malicious purposes [241].
- *Liveness Attack*: This attack delays the confirmation time of a specific transaction, allowing the attacker to continue building a private blockchain and incorporating blocks into the public blockchain [242].
- *Border Gateway Protocol (BGP) Hijacking Attack*: The attacker reroutes traffic to a mining pool they control, enabling them to steal cryptocurrencies [243].

In [244], further security concerns related to the Internet of Entities (IoE) are explored, specifically in the integration of mobile networks and the IoT to deliver services. The authors propose a Blockchain-based paradigm to enhance security within this environment. Similarly, Huang et al. huang2017decentralized present a decentralized solution utilizing blockchain for secure data exchange in the IoT domain. Similarly, Huang et al. [245] present a decentralized solution utilizing blockchain for secure data exchange in the IoT domain. Their proposed architecture leverages Ethereum and smart contracts to establish trust in the data sharing process.

Despite the proliferation of Blockchain-based applications in various domains such as finance, reputation systems, and the Internet of Things (IoT), there is currently limited research on Blockchain architectures specifically tailored for government applications. To the best of the authors' knowledge, there have been no proposed architectures that address citizen authentication through the combined use of smart cards and biometric technology. Notably, Estonia has emerged as a leader in Blockchain adoption within the government sector. Since 2012, Estonia has successfully implemented Blockchain systems in healthcare, judiciary, legislation, security, and commercial code systems. Furthermore, plans are underway to expand its usage to other areas, including medicine, cybersecurity, and embassy data. In addition, Estonia introduced the e-residency program in 2014, allowing non-Estonians to access various services provided by the government, as mentioned earlier [246].

In [247], the utilization of biometry is suggested for enhancing the security of digital identity models and its application in smart devices. These devices require a reliable means of uniquely identifying users to provide personalized services, and biometry offers a suitable technology for this purpose. By combining biometry with

Blockchain technology (BCT), the proposed approach aims to ensure the integrity of biometric features by storing them securely in the blockchain, thereby mitigating risks such as spoofing or related attacks. Furthermore, [248] highlights the numerous advantages that BCT can bring to both users and developers of biometric systems, including improved security, scalability, and privacy. Consequently, the integration of these two technologies complements each other in terms of enhancing overall security measures.

6.1.2 Authentication Methods

Document Authentication

The authenticity of documents is verified and validated using various methods. This subsection provides a brief review of the five most commonly employed document authentication techniques: changeable laser image, holograms, watermarking, one-way functions, and protocols.

Changeable Laser Image (CLI): CLI is a technique that involves using a laser to create a pattern on a document. This pattern is designed to change its appearance when viewed from different angles. The process of CLI includes laser drilling from various angles to achieve the desired image effect. Laser marking offers notable advantages such as durability and resistance to tampering. A notable example of CLI application can be found in the national identity document of Spain, where CLI is employed to authenticate the document by incorporating a changing image of the individual's photo and the date of issuance [249].

Watermarking methods: Techniques that involve concealing information within watermarks, entail the insertion of a message, whether hidden or visible, within a digital object (such as images, audio, video, text, or software) to authenticate the document's legitimacy. These watermarks can be categorized as follows: i) visual, which are hidden marks only detectable by the human eye through light reflection; ii) static software watermarking, encompassing data and code; and iii) dynamic software watermarking, encompassing data structures and the sequence of operations performed [250].

Holograms: Another approach to document authentication involves the use of holography, which encompasses various mechanical and visual methods. These

methods employ elements such as optical variable ink (OVI), orlas, guilloch 3D, microtext, kinegram, reliefs, tactile letters, ultraviolet and infrared inks, and OCR-B characters. In [251], holograms are described as three-dimensional images created by reproducing an interference pattern recorded with coherent light beams. These holographic images exhibit different figures through the manipulation of light, either through movement or reflection. In a recent study [252], the concept of holograms has been further advanced with the introduction of a digital holography (DH) system for authentication purposes. This system utilizes a charged-coupled device (CCD) camera and numerical 3D image reconstruction techniques to generate holographic representations.representations.

One-way functions These functions, associated with software methods, are a type of functions that are relatively easy to compute but exceedingly difficult to reverse. They take input data (or files) of varying lengths and generate a fixed-length string. Even a minor alteration, such as modifying a single character, bit, or file property, can result in a substantial change in the resulting string. This property enables the assurance of information integrity. Examples of these methods include hash algorithms (such as SHA-1, SHA-2, and SHA-3) and cryptographic hash functions (like MAC) [253].

Document authentication protocols: These protocols are employed to tackle security concerns in open networks, aiming to verify the legitimacy of the connecting nodes. They serve as the initial and crucial defense mechanism within a system of trusted and open networks [254]. The commonly utilized protocols include Secure Sockets Layer (TLS/SSL), the IP Authentication header (IPSEC), Secure Shell (SSH), and Kerberos.

In summary, when it comes to national identification document systems, the predominant approach for document authentication involves the utilization of mechanical and visual methods. Specifically, the combination of Changeable Laser Image (CLI) and various holographic techniques is commonly employed. However, as highlighted in [255], these methods are static and predetermined, rendering them susceptible to potential risks such as duplication or counterfeiting. Therefore, there are still security challenges to address in national identification document systems, particularly when integrating different authentication methods for documents, users, and transactions, which necessitates further consideration from an implementation standpoint.

User Authentication

These methods involve the verification and validation of the relationship between the document and its owner. In this section, the definition provided in [230] is adopted, and three common factors used for authentication are considered: i) something that is or done by the user, such as biometric and behavioral characteristics, ii) something that is known by the user, such as PINs, passwords, and one-time passwords (OTPs), and iii) something that is possessed by the user, such as Public Key Cryptography (PKC) technologies.

The factor of authentication based on *"something that is or done by the user"* primarily encompasses various biometric technologies. These technologies involve sensors and techniques that enable the capture of distinct biometric characteristics of individuals. In [256], the authors elaborate on the utilization of biometric technologies for automated personal recognition, which relies on two types of biometric characteristics. The first category pertains to biological traits such as fingerprint, face recognition, palm print, hand geometry, and iris recognition. The second category encompasses behavioral characteristics such as gait, signature, voice, and typing pattern.

The operational principles of a biometric system are demonstrated in [257]. While numerous biometric technologies exist today, not all of them have reached full maturity. Therefore, [258] focuses on describing the most commonly utilized biometric technologies.

Fingerprint recognition system: It is one of the most established and extensively developed methods. It is considered the oldest biometric technology. In [259], various types of fingerprint readers are described, including optical, capacitive, ultrasonic, and thermal readers, which are used to capture the fingerprint characteristics. The reader scans and identifies the fingerprint pattern, known as minutiae, along with additional features like body temperature. According to [257], minutiae encompass information about bifurcations, ridge endings, dots, core points, and delta points within the fingerprint. The fingerprint reader captures a significant number of minutiae, which are then compared with stored minutiae, producing a numerical result indicating the likelihood of a match between the captured and stored fingerprints. The system's reliability increases with a higher number of acquired minutiae. Fingerprint systems offer several advantages, including distinctiveness or uniqueness,

permanence, and performance in terms of accuracy, speed, and robustness [260]. However, they also have certain limitations, such as lack of universality when individuals do not possess fingerprints due to accidents, ageing, issues of acceptability, and vulnerability to circumvention through imitation using artifacts or substitutes.

Face recognition system: It is a biometric technology that focuses on identifying facial features and utilizes a digital camera to convert the user's facial image into digital data for authentication purposes [257]. Unlike other biometric methods, face recognition is typically a passive system as it does not require individuals to actively interact with a scanner. This system is capable of capturing face images from a distance using a video camera, and through the utilization of a face recognition algorithm, it processes the acquired data to detect, track, and ultimately recognize individuals. Face recognition involves the computerized recognition of personal identity based on geometric or statistical features extracted from facial images [261].

Iris recognition system: It is a type of biometric technology that focuses on analyzing the unique features present in the colored tissue ring of the iris [257]. Compared to other eye biometric technologies, iris scanning is considered to be less invasive as it utilizes a conventional camera element and does not require close contact between the user and the scanner. Furthermore, iris recognition has the potential for achieving higher matching performance than average. As a result, iris biometrics are particularly effective in identification mode [261].

Authentication based on the factor of *"something that is known by the user"* primarily revolves around the use of passwords or PINs that individuals are familiar with

A Personal Identification Number (PIN): A PIN is a numerical password utilized for user authentication in a system. Unlike being printed or embedded on the card, the PIN is manually entered by the user [262].

The password: A password is a word or combination of characters utilized by a user to verify their identity and must be kept confidential. Various techniques employed by attackers to acquire passwords include guessing, dictionary attacks, brute force, and rainbow tables [263].

One-time Password (OTP): OTP is an advancement in password authentication that reduces the risk of replay and brute force attacks. It offers a unique and valid password for each authentication session, requiring the generation of a new OTP for

subsequent access. This approach significantly limits the success rate of a *Man In the Middle* (MitM) attack.

Zero-Knowledge Proofs (ZKP): Another user authentication solution is Zero-Knowledge Proofs (ZKP). ZKP enables a user to prove the truth of a statement to a verifier without disclosing any sensitive information or procedures. Financial institutions often employ this method by randomly asking customers personalized questions that were previously established by the customers themselves [264].

The final authentication methods rely on the "something that is possessed by the user" factor, it is primarily associated with Public Key Infrastructure (PKI). PKI involves three entities: a client, a server, and a certification authority. The certification authority acts as a trusted entity that verifies the identities of the communicating parties and manages the validity of the certificates. Public Key Cryptography (PKC) plays a key role in this context and generates two keys [265]. One key, known as the private key, is securely stored, while the other key, known as the public key, is made publicly available. The private key is used for encryption, and the corresponding public key is used for decryption, ensuring non-repudiation [266]. PKC finds extensive application in broadcast authentication [267], where various technologies are employed for writing or reading information. These technologies are described below:

Barcode: Barcodes offer quick data capture, immediate integration of decoded data into systems, and cost-effective printing. There are two main types of barcodes: one-dimensional (Figure 6.1a) and two-dimensional (Figure 6.1b and 6.1c). One-dimensional barcodes consist of a series of bars and spaces designed to be scanned and read by a computer, but they have limited storage capacity, typically ranging from 20 to 30 digits. Examples of one-dimensional barcodes include EAN 13, EAN 8, CODE 128, DUN 14, and UPC 39. To ensure data integrity, barcode labels often include a control digit that prevents data loss during scanning, and any data failing the set point digit test is not entered into the system [268]. Two-dimensional barcodes can store more information compared to one-dimensional barcodes. For instance, PDF-417 can hold a maximum of 1,800 alphanumeric characters (ASCII) or 1,100 binary codes per symbol (each rectangle forming a "point cloud"). Another example is the QR Code (Quick Response Code), which is an ISO international standard (ISO/IEC18004). It utilizes four standardized encoding modes: numeric (7,089 characters), alphanumeric (4,296 characters), byte/binary (2,953 bytes), and kanji

(1,817 characters) for Japanese writing [269]. However, QR Codes can pose security risks as they can be manipulated to change encoded information, potentially enabling SQL injection, command injection, phishing and pharming attacks. Additionally, reader software vulnerabilities and social engineering attacks are also possible [270].

(a) (b) (c)

Fig. 6.1 Barcode technology evolution: a) Code 128. b) Code PDF417 and c) QR Code.

Smart card: A smart card consists of an embedded microprocessor and a memory, serving various purposes such as identification, data access, authentication, security key storage, and financial transactions [271]. When a smart card is inserted into a Card Acceptance Device (CAD), the metallic pads on the card make contact with the corresponding metallic pins on the CAD, enabling communication between the two. Upon insertion, smart cards are always reset, triggering the transmission of an "Answer-to-Reset" (ATR) message by the card. This message informs the CAD about the communication protocols and transaction processing rules associated with the card.

Radio Frequency Identification (RFID): RFID is a technology that enables proximity identification and data transactions using electromagnetic signals. Its effective range varies from 10 cm to several meters, depending on the type of RFID tag being used. This technology is widely employed for automated identification of objects and individuals, and it has greatly contributed to the advancement of ubiquitous or pervasive computing [272]. An RFID system consists of three primary components: a tag or transponder, which is typically affixed to the object or person being identified and serves as the RFID device; a tag reader or transceiver equipped with an antenna, which facilitates reading and writing of data on the RFID tags; and a data processing subsystem that utilizes the information obtained from the tag reader. The data processing subsystem is usually connected to a host system or an enterprise system [273]. The storage capacity of RFID tags is typically limited, ranging from 1 kilobyte (Mifare card) to up to 8 kilobytes [274].

Near Field Communication (NFC): NFC is a specific application of RFID technology that enables wireless communication and data exchange between devices in close proximity, typically within a distance of up to 10 cm. One of its key advantages

is its simplicity, as a transaction can be initiated simply by bringing an NFC-enabled device close to a reader, another NFC device, or an NFC-compliant transponder. NFC devices operate in three different modes: peer-to-peer mode, reader/writer mode, and card emulation mode. In peer-to-peer mode, two NFC devices establish communication and exchange data. A device operating in reader/writer mode can read or write various types of information, such as URLs, SMS messages, and phone numbers, and it can store data capacities of up to 32 KB (Type 4) [275]. Lastly, the card emulation mode allows an NFC device to emulate a contactless smartcard [272].

Utilizing the aforementioned technologies, it becomes feasible to authenticate individuals or documents. Although governments employ multiple methods to identify citizens and mitigate identity theft, the authentication of transactions still faces significant challenges.

6.1.3 Proposed Architecture for Biometric E-ID System with Blockchain

The proposed system offers enhanced security measures for both verifying and authenticating the document holder and the document itself. Furthermore, to reinforce the validation process of the document and the conducted transactions, blockchain technology is utilized. This integration incorporates existing security measures such as holograms, barcodes, and 3D images found in conventional identification documents. The Smart Card technology, equipped with a cryptographic chip, is employed to store all relevant bearer information. By implementing blockchain technology, the system ensures the recording and verification of transactions carried out by registered citizens in the electoral census. For further information on the transaction and block details of blockchain systems, refer to [276].

Blockchain Architecture

Figure 6.2 depicts the distributed and decentralized network architecture, encompassing all network nodes, the blockchain-based database, and government entities. The network functions as a private cloud, exclusively allowing participation from nodes situated in notarial and registry entities. Three types of nodes are present within the

network. Firstly, nodes located in register offices are responsible for issuing new or duplicate identity documents upon a citizen's personal request. These entities are exclusively authorized to generate digital certificates, comprising public and private keys, for each citizen using their unique PIN. Through this digital certificate, citizens gain the ability to sign digital documents, create accounts, conduct transactions, and authenticate themselves with relevant authorities. Secondly, nodes located in notary offices maintain records of individuals' civil status. Their role involves verifying the alignment between people's identities and their corresponding documents by providing testimonial evidence of authenticity.

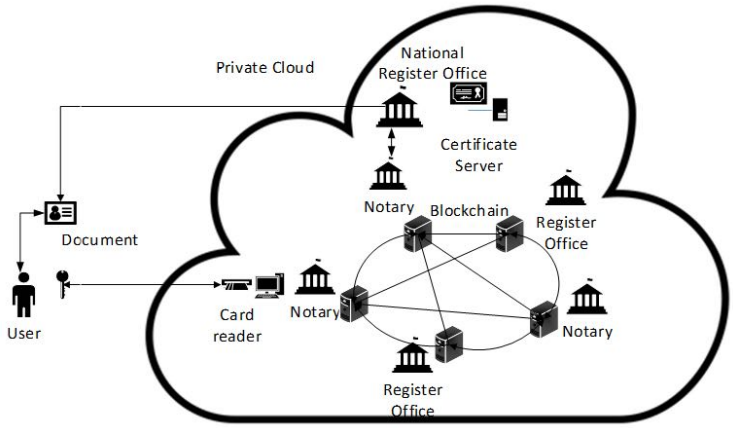


Fig. 6.2 Proposed Blockchain architecture for the e-ID System [276]

Blockchain Node Architecture

The block diagram presented in Figure 6.3 illustrates a node within the proposed network architecture. A citizen who possesses their own electronic DNI (Digital National Identity) utilizes a Match on Card (MoC) system to generate a security certificate. This MoC system enables local verification of the templates stored on the card, which include fingerprint and iris templates. The templates are read by a standard 32K NFC storage chip, adhering to the common criteria EAL 5, and allowing user authentication with the national government using generic software. Throughout the transaction, both the media on the card and the encryption provided by the Blockchain ensure the protection of data. All nodes within the network are aware of the transaction's existence, but only the receiver possesses knowledge of its content.

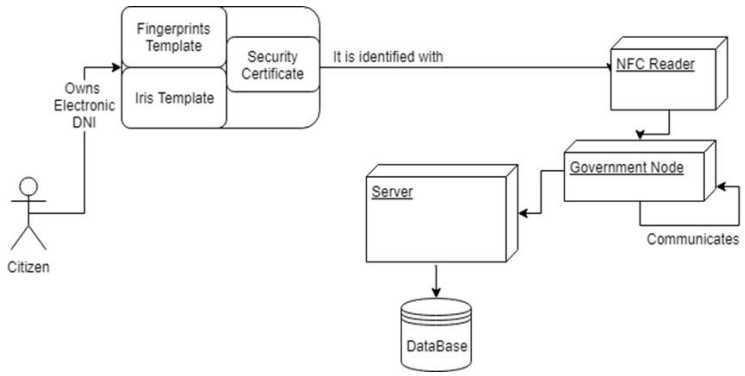


Fig. 6.3 Block diagram of a node in the proposed network architecture

Typically, the software employed for user identity validation exhibits security vulnerabilities, including potential NFC failures that may lead to identity phishing and the acceptance of mismatched fingerprints. To address the threat of user impersonation, our proposal incorporates a dual biometric user validation, alongside the encryption of public and private keys within the security certificate via NFC. Moreover, an extra layer of security is ensured by storing transactions using Blockchain technology in a distributed system, which guarantees traceability, security, and scalability.

The solution put forth adopts a private model wherein all nodes are owned by a single organization, namely the national government. This approach ensures that the system operates within strict boundaries, delivering high efficiency and enabling centralized information management. By integrating this system with the document, the likelihood of information compromise or impersonation is significantly minimized.

Deployment Model

To illustrate the physical relationships between hardware and software components, the software deployment diagram is presented in Figure 6.3. The transaction process carried out by a citizen is described as follows:

1. The citizen visits a government office, such as the National Register office, local register office, or notary's office. They access an authorized computer called *Webserver identification*, which is equipped with a card reader and a biometric reader.

2. The citizen utilizes their card and biometric feature (iris) to authenticate themselves.

3. Upon successful identification, the citizen's information is automatically displayed on a form. Subsequently, the citizen can proceed with a transaction by selecting an option from a drop-down list.

4. The *Identification webserver* is connected to a database that stores citizen data, including digital signatures, fingerprint templates, and iris information. This enables the verification of the citizen's authenticity.

5. Finally, the *Identification webserver* will present a message on the screen confirming the successful validation of the user.

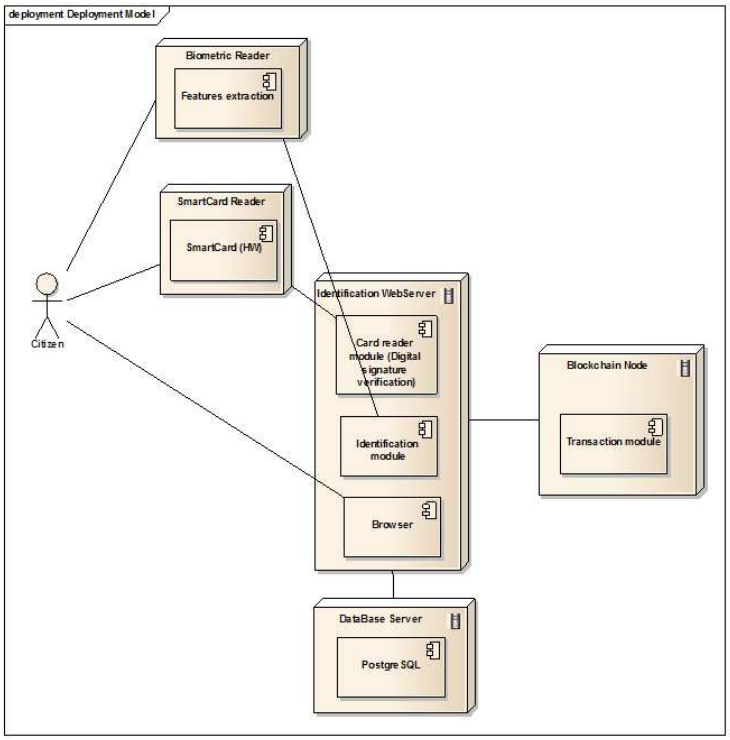


Fig. 6.4 Deployment model

Transactions

Once the system has successfully validated and authenticated a citizen, they can proceed to perform a transaction by selecting one of the predefined options from

a drop-down list. The available transactions may include actions related to the citizen's personal information, such as updating their civil status, registering a child, requesting a duplicate document, and other relevant operations, as depicted in Figure 6.5.

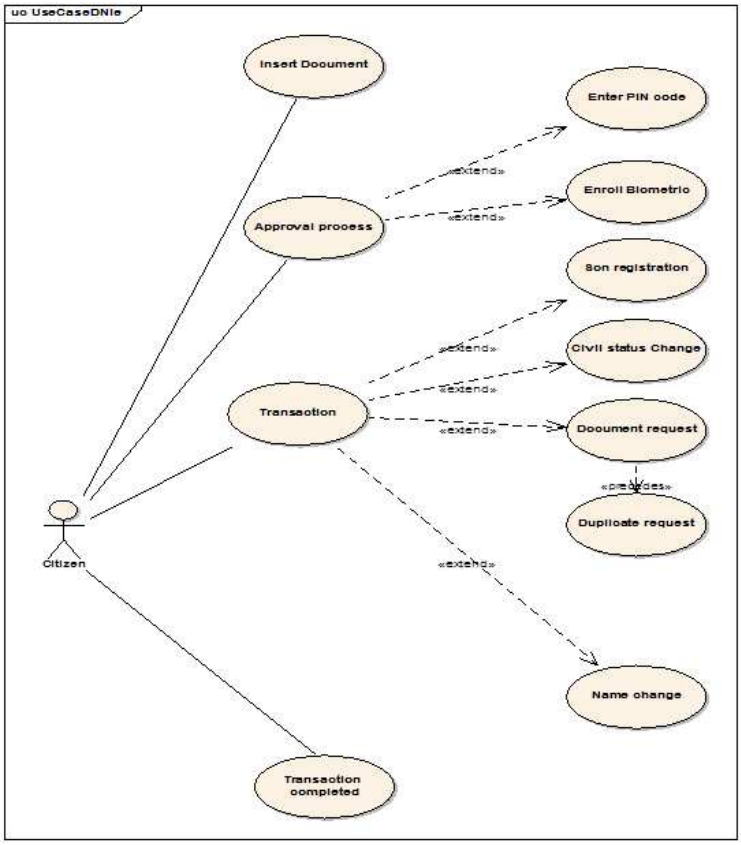


Fig. 6.5 Use case diagram

System Implementation

Each time a user initiates a transaction, it undergoes a validation process to ensure its correctness and adherence to the required format. Once verified, the transaction is signed using the user's digital certificate. Subsequently, the transaction, along with its relevant information such as the Transaction ID and Issuer, is recorded and stored. These transactions remain in a pending state until a specified number, such as one thousand (1000), are accumulated, at which point a new block is created.

The exact number of transactions per block can be configured based on performance considerations. Upon the creation of a new block, the process of determining

which node will be responsible for mining it begins, following the consensus algorithm proposed in the following subsection. Finally, once a new block is successfully mined, it is incorporated into the Blockchain, ensuring the secure and immutable storage of the recorded transactions.

6.1.4 Consensus Algorithms

These algorithms play a crucial role in the functioning of the system as they are responsible for adding new blocks to the Blockchain and maintaining control over it. The nodes within the Blockchain network are tasked with meeting certain requirements imposed by the system, typically involving the generation of a hash function output with a specified number of leading zeros. The choice of consensus algorithm is important as it can significantly impact power consumption. For instance, the proof of work algorithm can be energy-intensive, making it less suitable for private Blockchains. Alternatively, the proof of stake algorithm offers an alternative approach, or it is also possible to design a custom algorithm tailored to the specific needs and environment of the Blockchain system.

Proof of Work (PoW) Algorithm

The proof of work algorithm, which was initially introduced with the creation of Bitcoin in 2009, is widely recognized and known. It involves the process of finding a "magical number" called a nonce, which grants the right to add a new block to the Blockchain and receive a reward. The nonce value is incremented from zero until a number is found that produces a hash function output with the required number of leading zeros, as specified by the system. This process is repeated multiple times by calculating the SHA256 hash function and checking if the condition is satisfied. In essence, mining refers to the repetitive hashing of the block header, adjusting one parameter at a time, until the resulting hash meets a predetermined target. The only way to find a specific solution to the problem is through trial and error, similar to a brute force attack.

Each node in the network competes to find a solution, and when a node successfully finds a solution, it is rewarded. This process involves every node attempting to solve a mathematical operation, and once a solution is found, other nodes stop their search and verify the proposed solution. The solution is considered valid if at

least 51% or more of the nodes approve it. This process is known as mining, and the individuals or nodes involved in solving the mathematical operation are referred to as miners.

Mining plays a crucial role in securing the Blockchain system without relying on a central authority. It prevents anyone from creating a block at will, and it also serves to generate new cryptocurrency within the system through the rewards received by miners. The creation of cryptocurrency acts as an incentive system that helps maintain the security of the decentralized ledger

The difficulty target in the Bitcoin Blockchain is adjusted approximately every 2016 blocks in order to maintain an average block time of 10 minutes. This adjustment is based on the collective computing power of all the participating nodes in the network. However, this mechanism poses significant environmental concerns. Due to the high value of bitcoins and the potential for substantial rewards, a large number of individuals are motivated to engage in mining activities. As a result, the difficulty target is set at a high level, requiring a significant amount of computational power to solve. This high level of difficulty translates into a considerable consumption of electricity and energy resources, contributing to environmental challenges associated with Bitcoin mining.

Once a solution is found, it is broadcasted to all nodes in the blockchain network to add the block to the public ledger. However, due to the decentralized nature of the network, it is possible that not all nodes receive the validated block at the same time. In some cases, another node may find a solution and start broadcasting it before everyone receives the initially validated block. This situation leads to what is known as the forking problem, where different versions of the ledger exist with different chains of blocks.

To address this problem, each node continues with the block it received first. Over time, one chain will become longer than the others due to factors such as more nodes mining on that chain. In such cases, the rule is simple: the longest chain is considered the valid one. Consequently, when a forking problem occurs, the process continues, and eventually, the longest chain is accepted by all nodes. While the youngest block in a chain can be subject to variation during this process, the difficulty of adding a new block ensures the overall stabilization of the blockchain.

Security Problems in PoW

Blockchain technology is widely regarded as highly secure due to the inherent difficulty of falsifying a block or altering the ledger. Nonetheless, certain security concerns have emerged, prompting the development of various proof-based algorithms aimed at addressing these issues.

For instance, consider the scenario where the contents of an older block are modified. In such a case, the validity of the block's hash is compromised, as the previously found Nonce value was suitable for the original transactions. As observed, even a slight alteration in the input of a hash function yields significant changes in the output. Consequently, all subsequent blocks linked to the invalidated block through the previous hash block field in the header also become invalid. The input for the hash function is derived from the header, rendering the previously determined Nonce value ineffective for all subsequent blocks. This means that modifying a portion of the Blockchain using the hash function renders all subsequent blocks false, necessitating the rediscovery of Nonce values for the affected blocks. Due to the demanding nature of the puzzle and the rule that the longest chain is considered valid, tampering with the Blockchain may seem impossible. However, if an individual possesses an extensive array of modern hardware resources that surpasses the combined computing power of all network nodes, they could potentially falsify the entire Blockchain. This attack, known as the Double spending attack or the 51% attack, references the minimum computing power required to carry out such an exploit [234].

Currently, the likelihood of a single node executing such an attack is extremely low. However, the emergence of mining pools introduces the possibility of carrying out such attacks. To address this concern, extensive research has been conducted [277]. In a mining pool, the rewards are distributed evenly among all participants. In order to disrupt pool mining, Miller et al. [278] proposed an enhanced version of the Proof of Work (PoW) algorithm known as non-outsourceable puzzles. This consensus algorithm aims to discourage mining pools by introducing a mechanism that allows a miner within the pool to claim all the rewards without exerting any individual effort.

In the Ethereum cryptocurrency, an alternative technique known as GHOST (Greedy Heaviest Observed SubTree) is employed. This consensus algorithm mandates that mining nodes include the headers of recently orphaned blocks, referred to as uncles, in the block header they intend to validate. Orphaned blocks are blocks

that have been added to parallel branches of the primary Blockchain. In the context of Bitcoin, an uncle is considered an orphaned block because it does not reside on the longest chain. Unlike the traditional approach of selecting the longest chain, the GHOST strategy selects the chain with the highest number of PoW contributions as the valid chain. Ethereum incentivizes miners to include a list of uncles when validating a new block. This technique has two primary effects:

- It reduces the incentive for centralization by providing minimal rewards to miners who produce obsolete or orphaned blocks. Since these miners are not part of a large group and receive notifications about other blocks later due to network propagation delays, their rewards are diminished.
- It enhances the security of the chain by augmenting the cumulative work on the main chain. As a result, less work is wasted on alternative branches in favor of the main branch, leading to a more secure Blockchain.

Proof of Stake (PoS) Algorithm

In this system, every node in the network must demonstrate its ownership of a specific portion of the circulating tokens in order to participate in the block validation process. The consensus algorithm then selects one of the nodes with the largest stake to validate a new block. In a PoS context, the probability of a node being chosen to validate a new block is proportional to its stake or percentage of the circulating tokens (currency or shares) it possesses. If a node owns x quantity out of a total of y coins, its chance of appending the next block is x/y. This selection process is carried out in a pseudorandom manner to prevent a node from predicting when it will be chosen to validate the next block. However, certain additional parameters, such as the duration of token ownership, may be considered, allowing the "wealthiest" nodes to be selected more frequently. It is important to note that the validation of a block does not directly result in monetary compensation; instead, holding a certain amount of cryptocurrency generates a reward similar to earning interest.

PoS offers two notable advantages over PoW. Firstly, it is a more energy-efficient mechanism as it requires significantly less energy consumption compared to PoW, which relies on extensive cryptographic calculations for block validation. In contrast, PoS operates based on the ownership of a certain percentage of circulating tokens, eliminating the need for intensive computational power. Secondly, PoS makes the

51% attack more challenging. In a PoS system, executing a 51% attack requires controlling more than half of the circulating tokens, which is typically a more expensive and difficult task compared to controlling 51% of the computing power in a PoW system. As a result, PoS provides increased resistance against such attacks.

Security Problem in PoS

PoS algorithms also come with certain drawbacks and security concerns. One prominent issue is known as *tNothing-at-Stake*. In pure PoS algorithms, nodes lack the incentive to vote for the most likely legitimate chain, such as the longest chain in the case of Bitcoin. When multiple potential chains exist, nodes aim to maximize their chances of receiving rewards by evenly allocating their stake and simultaneously voting for the latest blocks in the potential chains. Unlike PoW, where mining on multiple chains requires energy expenditure and monetary costs for miners, PoS allows mining on multiple chains at no cost. This situation creates an opportunity for miners to prioritize their personal interests, potentially facilitating double-spending attacks. An attacker could initiate a fork from a block preceding a transaction and redirect the funds to themselves, even with just a 1% stake, as their fork would prevail due to other miners mining on both chains simultaneously.

In a pure PoS system, the selection of the block appender is based solely on the stake owned by the miner. This means that miners with a larger stake have a higher probability of being chosen. However, this approach could result in undesirable centralization, as the richest member would always have the advantage. To address this issue, alternative methods have been suggested in the literature [279, 280] to mitigate the concentration of power and promote a more decentralized system.

Satoshi Consensus Algorithm

The Satoshi consensus algorithm, as described in Bentov et al.[281], utilizes the state of the block to determine the miner responsible for appending the next block. A Satoshi, the smallest unit of currency in bitcoin, serves as the basis for this selection process. To determine the block appender, a random index number is generated within the range of 0 to the total number of satoshis. Subsequently, the transactions associated with this specific satoshi are identified, and the current owner of that satoshi is designated as the miner responsible for appending the next block.

Proof of Luck (PoL) Algorithm

The proof of luck algorithm operates by distributing transactions to participating nodes, who then construct a new block by committing these transactions. This algorithm incorporates Trusted Execution Environments (TEE), which are secure hardware environments like Intel SGX-enabled CPUs. Within a TEE, a proof of luck is generated, and the participant node uses this proof to extend its chain and subsequently broadcast it. The new block includes the hash of the previous block, the data comprising new transactions, and the proof of luck. This proof of luck involves computing a numeric score, known as luck, for a given blockchain. The luck score is obtained by summing a predetermined set of values from each block. The chain with the highest luck score is selected as the preferred chain for further validation and continuation [282].

Consensus Algorithm for Private Blockchain

The Proof of Luck (PoL) algorithm incorporates Trusted Execution Environments (TEE) to ensure the proper execution of desired code and prevent potential attacks on the Blockchain. However, considering that the proposed private Blockchain in this section operates within a trusted environment where nodes are inherently trusted, the use of TEE becomes unnecessary, eliminating the need for additional hardware requirements. As a result, in the following subsection a novel consensus algorithm called the Tournament Consensus Algorithm is proposed.

Tournament Consensus Algorithm (TCA)

When a node seeks to join the blockchain, a request is sent to all nodes to obtain authorization. Upon confirmation as an authorized node, it will be granted access to the blockchain, the ability to create transactions, and the opportunity to attempt to append the next blocks. At random intervals, each connected node is prompted to select a number between 0 and 1, typically within a defined time frame (e.g., 15 seconds as suggested by the authors of the PoL algorithm, in comparison to Ethereum). Subsequently, every node broadcasts their chosen number and awaits receipt of the random numbers from other nodes. Upon receiving all the random numbers, a node selects the highest value and sends a "winner vote" to the corresponding node. Once a node receives "winner votes" equal to the number of connected nodes, it is designated as the winner and gains the privilege to add the next block. The winning

node then mines the block using a challenge that is considerably easier compared to the proof of work consensus algorithm, thereby enabling any standard computer to quickly find the solution. This approach effectively avoids time wastage and energy consumption.

TCA presents a straightforward mechanism for determining the mining node. By leveraging the favorable attributes of the PoL algorithm, it is achieved a improved response time, as illustrated in Figure 6.6, while effectively mitigating high-cost attacks on individual Trusted Execution Environments (TEE). Furthermore, the process of sending random numbers and executing the consensus algorithm can be initiated with configurable values.

6.1.5 System Performance Results

Several tests were conducted on the network to assess the effectiveness of the proposed consensus algorithm, the Tournament Consensus Algorithm. These tests involved measuring the mining and broadcasting time, cache memory usage, and CPU percentage. A specific transaction, as outlined in Table 6.1, was used as part of these tests.

ID Number	101
Serial number	1
Class	Notary
Type	Civil status
Information	Married

Table 6.1 Transaction example

The lab computers utilized for conducting various tests shared the following common specifications:

- *Processor*: Intel Core i5 750 with a clock speed of 2.66 GHz and 4 cores.
- *RAM Memory*: 4 GB.
- *HDD*: 512 GB.
- *Network Card*: Intel 82574L Gigabit Ethernet NIC.
- *Operating System*: Linux Ubuntu 18.04.

The mining process time for a single block and the subsequent broadcasting process of that block to all nodes in the Blockchain system is illustrated in Figure 6.6. The test was conducted using a lab computer, with the number of transactions being varied to obtain an estimated time. As indicated in the figure, mining a block containing one thousand (1000) transactions took approximately five hundred and thirty (530) milliseconds.

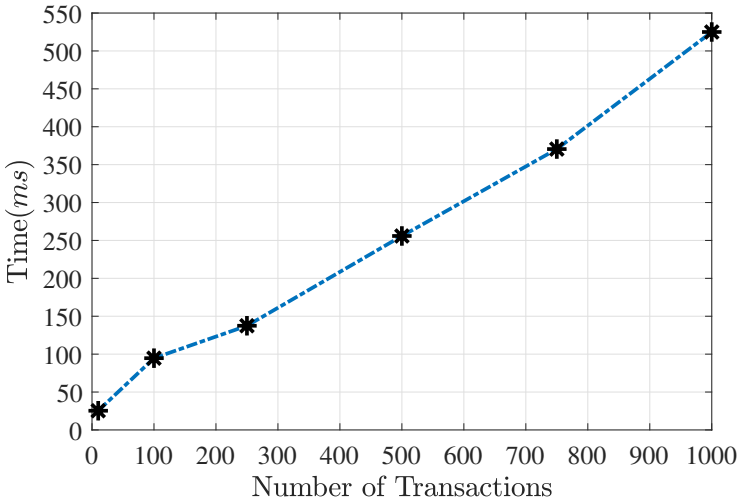


Fig. 6.6 Time for mining and broadcasting one block versus the number of transactions.

To determine the required cache memory for storing multiple transactions until the corresponding block is finalized, a test was conducted. The results indicate that with one thousand (1000) transactions, the approximate RAM usage is six hundred (600) Megabytes, as shown in Figure 6.7.

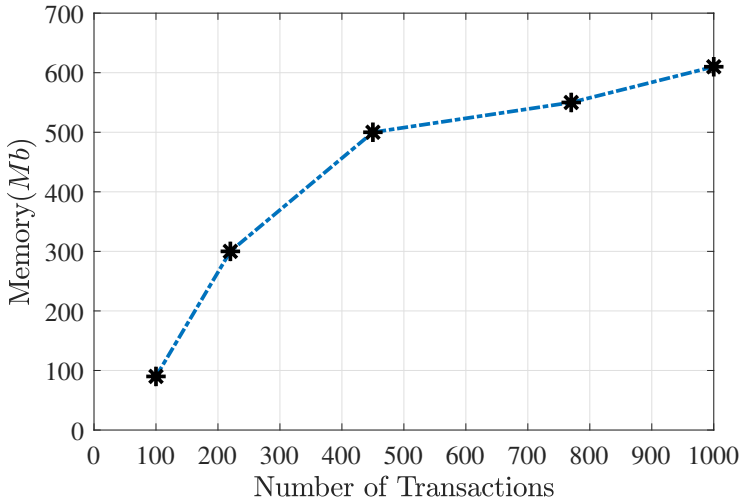


Fig. 6.7 Used memory for broadcasting a number of transactions

The CPU usage of a miner computer during the broadcasting process is illustrated in Figure 6.8. Initially, as the number of transactions increases from one hundred to two hundred (100-200), the memory consumption rises rapidly. However, after reaching two hundred ten transactions, the memory consumption continues to increase, but at a slower rate. Consequently, when broadcasting one thousand (1000) transactions, the CPU usage reaches approximately 43%.

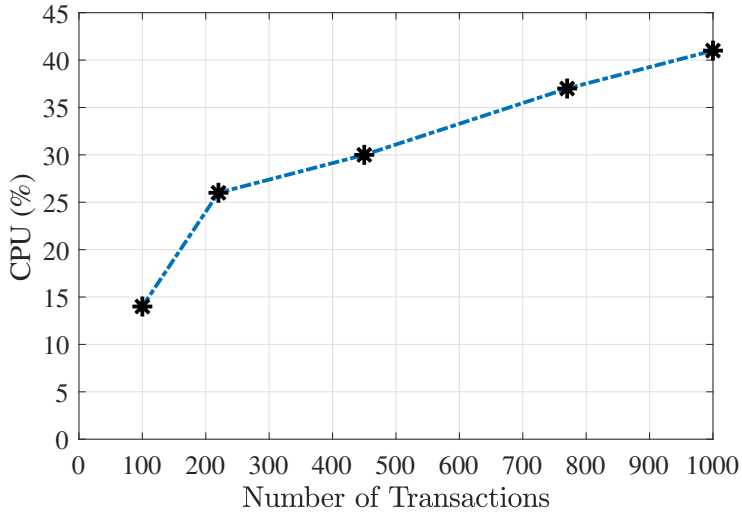


Fig. 6.8 CPU usage for creating and broadcasting a number of transactions.

Different strategies can be applied to address the considerable CPU usage percentage. For instance, if the number of users is not excessively high, one approach

is to adjust the number of transactions by setting a limit on the CPU usage. For instance, based on Figure 6.8, it can be observed that with a CPU usage limit of 30%, approximately four hundred fifty (450) transactions could be managed effectively.

In cases of special events or occasions when a significantly larger number of transactions are expected, such as during popular elections, it might be necessary to maintain the number of transactions at a higher level, such as one thousand (1000). However, in such situations, dedicated machines with better hardware specifications can be employed specifically for the purpose of creating and broadcasting transactions when a block is mined. This ensures that the increased transaction volume can be handled efficiently and effectively without overburdening the overall system

6.2 Air Pollution Monitoring over Blockchain Technologies

The previous section describes an approach towards private blockchain technologies, where the computation process is controlled by governmental entities. This blockchain model, while energy-efficient, relies on the network of authorized nodes being adequately protected to maintain data integrity. It would be significant to accumulate air pollution data, relayed in real-time via sensors, and subsequently store them in a blockchain, also known as Distributed Ledger Technology (DLT), to ensure their immutability.

This chapter features the design and development of a software framework enabling the integration of an Ethereum-based blockchain with the IoT framework presented previously in Chapter 4. The recommended solution is presented as an alternative method for incorporating a broad range of IoT devices, circumventing the reliance on centralized intermediaries and third-party services, thereby addressing the weaknesses of a private blockchain.

6.2.1 Introduction

Over the past decade, the emergence of novel technologies has paved the way for a dynamic interconnection between the digital and physical realms, giving rise to what is now known as the Internet of Things (IoT) ecosystems. These systems, defined

by their ability to link a myriad of devices into a cohesive, communicative network, have been integral in the evolution of our technological landscape.

In parallel, blockchain technology has experienced an equally, if not more, rapid evolution. This has been spurred on by its diversified use-cases, ranging from its foundational role in cryptocurrencies [231] to its significance in non-fungible tokens (NFTs) - a cornerstone for digital collectibles and markets [283]. Another pivotal application of blockchain lies in the realm of smart contracts. These are predefined, algorithmic processes that can be decentralized and securely inscribed on the blockchain. The procedures prescribe specific responses to new information inputs, such as incoming transactions [284].

Interestingly, the convergence of IoT and blockchain technologies can be leveraged to harness the benefits of both spheres. With this fusion, IoT devices gain the capability to independently authenticate transactions and record them on the blockchain. This process eliminates the necessity for third-party intermediation. Global perceptions of monitoring systems and their produced data vary, with a segment of the population questioning their reliability and credibility. The use of this type of technoligies could enhance the user confidence in environmental monitoring. The implementation of a blockchain serves this purpose, offering a robust method to validate and preserve the authenticity of the data. The synergistic combination of IoT and blockchain provides a formidable solution that significantly benefits areas such as environmental monitoring and citizen satisfaction.

6.2.2 Background

Several approaches to integrate IoT with blockchain have been investigated. In [285], the primary challenges associated with the the merging of these two technologies, such as storage issues, are explored. The resource constraints and system heterogeneity may prove inadequate for the management of a secure blockchain. Concerning throughput, the diversity in communication (see Chapter 4) makes achieving consensus costly. Thus, the use of intermediaries is often proposed to facilitate interaction with the blockchain, relieving devices from consensus mechanisms and storage responsibilities. This model is also proposed in [286], which outlines a decentralized architecture requiring delegation to intermediary nodes for interaction with the

blockchain. In [142], the focus is directly on air pollution data, using LoRa and a relay device for blockchain interaction.

Another alternative involves the use of blockchain schemes that do not employ mining but depend on a Hash Directed Acyclic Graph (DAG), more commonly known as Tangle, to record transactions. This mechanism forms the basis of IOTA, an open-source cryptocurrency developed for IoT with the aim of eliminating intermediaries, with a detailed explanation provided in [287]. However, this system places a burden on IoT devices in the calculation of PoW, resulting in energy costs and fees to secure the network.

Drawing from the literature, the most common solution can be summarized as shown in Figure 6.9. In this architecture, the data flow generated by sensors is funneled through an IoT Hub, which manages data aggregation, device administration, and in some cases, data transcoding (LoRaWAN), and centralized storage. An IoT-Blockchain intermediary is introduced, fully functional within the blockchain (i.e., a node within the peer-to-peer network with the capability to store, verify, and update data).

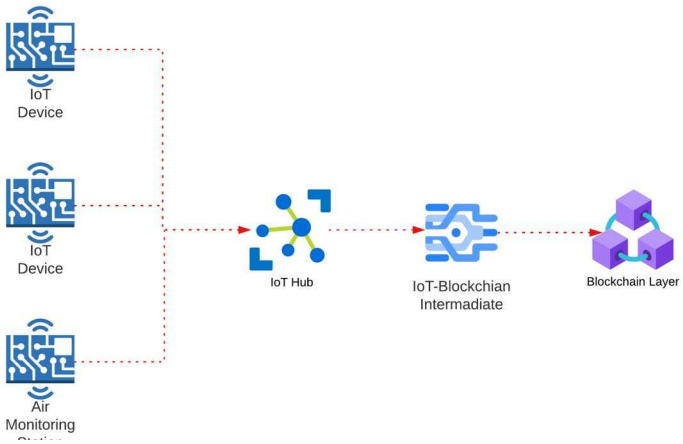


Fig. 6.9 Common architecture to merge IoT Solutions with Blockchain.

Numerous such models have been applied to air pollution monitoring in the literature. In [288], an IoT prototype is introduced to determine the air quality index through neural networks and proposes an architecture for data storage in blockchain. However, the details regarding the usage and application of blockchain are not presented. The authors of [289] propose a prototype architecture similar to the one described in Figure 6.9 but emphasize transaction and storage costs

on a public blockchain like Ethereum. They suggest a cost reduction by storing data on a distributed file system. This system, upon storing data, returns a hash, which due to its smaller size, is stored in the public blockchain. Although this architecture innovatively decentralizes storage, such solutions still present challenges in reliability.

Other authors in [290]propose a prototype with an intermediary-based architecture. However, they do not clearly define the type of blockchain used and the amount of data the transactions correspond to. Another alternative is presented in [291] where the architecture employs 5G transmissions for the interaction between the intermediary and the blockchain, but the overall architecture remains consistent with the previous models.

Lastly, in [292], a deployed implementation is presented using a station that generates measurements every 3 5 minutes. In this study, a distributed file system is employed once again for storage, combined with a centralized database, and Ethereum as the public blockchain. Under these conditions, they propose data storage optimization to reduce the transaction fees required to process transactions on Ethereum.

6.2.3 Proposed Architecture

Referring to the deployment architecture detailed in Chapter 4 (Figure 4.2), this architecture includes distinct elements specially designed for interaction with the Internet of Things (IoT) devices. These elements involve nodes for both storage and the retrieval of a centralized database, along with other nodes for comprehensive data analysis and processing.

This architecture, when juxtaposed with the more generic architecture meant for IoT and blockchain integration showcased in Figure 6.9, introduces these components under the umbrella term *IoT Hub*. The amalgamation of these elements results in a holistic system that facilitates data aggregation, effective device management, data transcoding (primarily via the LoRa Gateway), and safe data storage.

This proposal seeks to address specific vulnerabilities of the existing architectural framework, especially those associated with data integrity and potential manipulation. To mitigate these risks, the innovative solution incorporates blockchain technology that endows the system with the ability to verify the immutability of data from

a user's perspective. The comprehensive design of this proposed architecture is depicted in Figure 6.10 and consists of five layers.

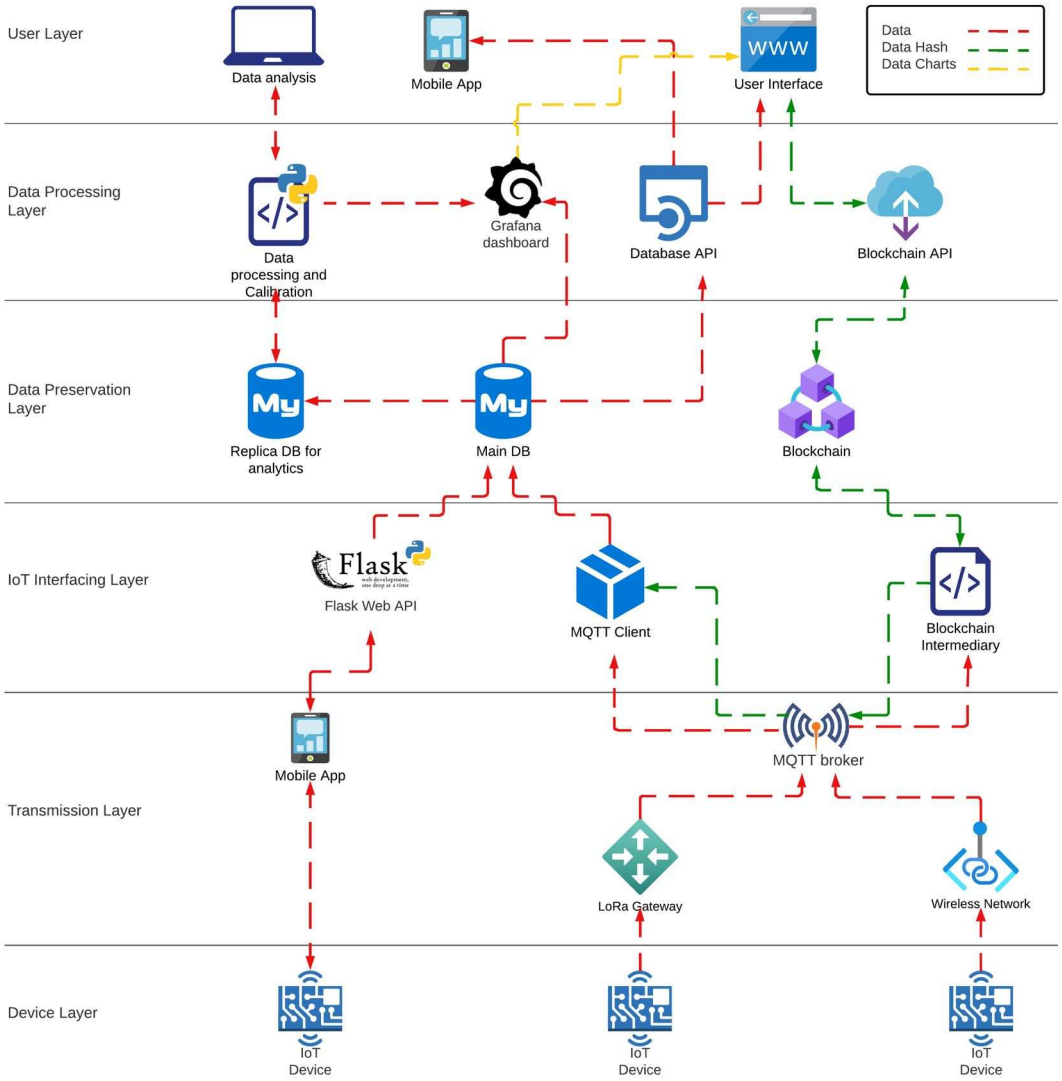


Fig. 6.10 Complete proposed architecture, IoT air monitoring with Blockchain.

Device Layer

At this stage, digitization of environmental variables, namely, temperature, humidity, and atmospheric pressure, is performed. Additionally, readings related to pollution due to particulate matter (specifically PM_{10} and $PM_{2.5}$) as well as GPS positioning are also digitized. The whole process of measurement acquisition and digitization

is controlled by the IoT device presented in Chapter 4. This device is responsible for data collection, formatting in accordance with the employed data transmission technology, and ultimately, transmitting the information to the subsequent layers in the architecture.

Transmission Layer

This layer consolidates various wireless transmission technologies, categorized according to the deployment type (refer to Chapter 4 for further details). The architecture embodies the requisite nodes for transmission through participatory sensing, facilitated by a *Mobile app*. This application acts as a bridge between the sensor, communicating via Bluetooth, and the succeeding layer that uses a mobile cellular network.

In the context of Low Power Wide Area Network (LPWAN) technologies, the model employs *LoRa Gateways*, capable of converting LoRa frames into MQTT messages over TCP/IP packets . Finally, in the *Wireless Network* category, all the wireless technologies compatible with TCP/IP protocols, such as WiFi, LTE, and others, are depicted. All transmitted data, barring the cases when participatory sensing is implemented, are encapsulated as MQTT messages. Communication with the IoT interface layer is enabled through the use of an *MQTT broker*.

IoT Interfacing Layer

This level embodies various nodes that gather and interpret data originating from IoT devices. Depending on the employed format, the decoded data is subsequently stored in the data preservation layer. The proposed architecture integrates three distinct nodes: a Flask server that fetches data from a mobile application via a RESTful API, and an MQTT client that subscribes to diverse MQTT topics generated by the IoT devices. Following the decoding of data, which is transmitted in binary format, these nodes proceed to deposit the data into a database.

Additionally, a blockchain intermediary node performs transactions that ensure data integrity within the database. It achieves this by generating a hash of the data received hourly from each device, which is then stored in the blockchain. Concurrently, it transmits via MQTT the metadata utilized in the hash generation for

its preservation in the database. A more detailed description of the hash generation process is presented later in the next subsection.

Data Preservation Layer

The present layer consolidates various databases deployed within the architecture, storing data derived from the measurements alongside the corresponding meta-information. This meta-information plays a crucial role in the creation of a unique hash that finds its place in a public blockchain.

The architecture introduces a parallel database. This database, designed with an identical structure to the primary one, functions as a sandbox for data analysis and experimentation. The utility of this parallel system is the ability to maintain the integrity of the original data during experimental trials, effectively providing a fail-safe mechanism for the proposed architecture.

At the same juncture, the public blockchain forms an integral part of this layer. The function of the blockchain extends beyond merely storing the hash of the raw data that it receives via MQTT in a specific time window. It preserves the hash of both the raw data and the associated meta-information, establishing a powerful, tamper-proof record.

Data Processing Layer

In this layer, the processing and analysis of the collected data, such as calibration and data visualization, are performed, facilitating the understanding of a pollutant's behavior in a smart city context. Each node within this layer interacts with the main database to retrieve data, ensuring no alteration in the data values during this process. Should any of these processes necessitate modifications, these changes are relegated to the parallel database, maintaining the integrity of the original data.

For the visualization component, the architecture employs Grafana Dashboard, an interactive visualization platform. This tool generates various graphical representations that can subsequently be incorporated into a user interface, thereby enhancing user interaction and understanding of the data trends.

Key to the verification of data immutability within this architecture are two central nodes: the Database API and the Blockchain API. The Database API allows for data

access and retrieval of essential values necessary for data set verification. Through a database query, this node facilitates the retrieval of the data and its sequence for calculating the hash that is stored within the blockchain. Upon the recalculation of the hash, the Blockchain API node is then utilized to verify whether the recalculated hash corresponds to the hash stored within the blockchain. This meticulous process ensures the rigorous verification of data immutability, significantly enhancing the reliability of the system.

User Layer

This final layer encapsulates the diverse interaction modalities with the architecture, catering to two distinct types of users: advanced and regular users. Advanced users (*Data Analysis* are granted access to both the parallel database and the data processing and calibration node, facilitating more in-depth data examination and manipulation.

On the other hand, regular users interact with the architecture primarily for data visualization purposes. They can access the data through a *Mobile App* interface or via a web interface (*user Interface*, both of which are designed to be intuitive and user-friendly. Moreover, the last one interface has access to various APIs, enabling regular users to verify the integrity of the data independently. This multi-tier user access system ensures that the architecture is adaptable to a wide range of user competencies, thus increasing its usability and effectiveness.

6.2.4 Blockchain Verification

The process of validating data through a public blockchain introduces certain limitations related to the volume of data stored within the blockchain. As highlighted in [292], an increase in stored data corresponds to a surge in transaction costs. Consequently, a more efficient approach involves storing a hash in the blockchain that symbolically represents the stored data. However, this approach stipulates the ability to reconstruct the stored hash during verification—a non-trivial process, as any alterations in the data sequence result in a differing hash from that recorded in the blockchain.

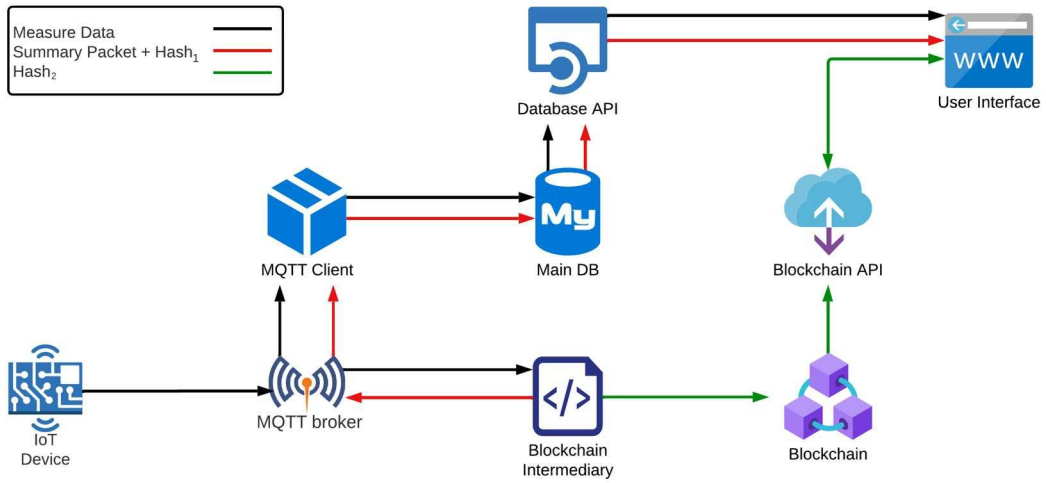


Fig. 6.11 Simplified blockchain architecture process.

Under this constraint, a specific schema for hash generation is devised (refer to Figure 6.11). In the initial phase, each IoT device generates data in binary format to optimize device resources. Accordingly, all measurements are encoded into binary and transmitted as positive integers, following the format detailed in Figure 4.11. The determination of this encoding is based on the resolutions and range of measurements of each sensor, as is shown in Table 6.2

Sensor Type	Byte Length	Conversion formula	Range	Resolution	Units
'T'	2	(measure + 40)*100	-40 to 80	0.01	°C
'H'	2	measure * 100	0 to 100	0.01	%RH
'P'	2	measure	0 to 10000	1	$\mu g/m^3$
'A'	4	measure * 10000	300 to 1100	0.001	hPa
'G'	4	(measure + 180) * 1000000	-180 to 180	0.000001	Degrees

Table 6.2 Binary coding of the measurements made by the IoT monitoring station

Prior to transmission, each measurement is stored in a buffer to minimize the amount of data used for header transmission. As such, each generated MQTT message comprises multiple measurements, determined by a predefined period that sets the transmission buffer's length, and the date is transmitted when the buffer is full. The use of a buffer presents an additional benefit, it reduces the number of write operations on each device's backup SD, thereby prolonging its lifespan.

Despite optimizing transmission, this format doesn't inherently facilitate the easy identification of each transmitted MQTT message and its arrival sequence at the hash generator node. It necessitates identifying when each packet is transmitted and which device is responsible for the measurement. Therefore, a proposal is made to send the data in JSON format, incorporating the following attributes:

```
{
    "block_ts": integer,
    "board_id": integer,
    "data": string
}
```

- *block_ts*: Signifies the transmission time of the MQTT message by the IoT device, represented using Unix timestamp encoding.
- *board_id*: Acts as the identifier for the IoT device, represented as a positive integer.
- *data*: Carries the binary data of the measurements, encoded as ASCII characters in base64 format.

The intermediary blockchain node receives each MQTT data packet and organizes these messages into one-hour blocks. Within each one-hour segment, a hash of the received data ($Hash_1$) is computed. Subsequently, a new JSON message is formulated using this information, containing the following attributes:

```
{
    "board_id": integer,
    "created": string <string timestamp>
    "hash": string <hash_1 sha256>
    "ip": string <board IP address>
    "timestamps": [<array of block_ts>]
}
```

- *board_id*: The identification number associated with the device responsible for generating the measurements.

- *created*: The date corresponding to the timestamp of the initial packet within the grouping.
- *hash*: The computed hash of the packet group received during the one-hour interval.
- *ip*: The IP address of the IoT device transmitting the measurements.
- *timestamp*: Arrays presenting the sequence of timestamp labels of the messages in the group.

Upon this JSON file, an additional hash ($Hash_2$) is generated, which is then stored in the blockchain via the intermediary node. The storage implementation is executed on the Ethereum network, which operates on a Proof of Stake (PoS) protocol, thereby reducing its energy consumption compared to other public blockchains that rely on Proof of Work (PoW) which is more demanding and energy-inefficient [22]. Additionally, Ethereum is one of the highest market capitalization blockchains, which indicates a high degree of interest and facilitates mitigating centralization issues (see 6.1.4.

Users who wish to validate the integrity of data can initiate this process via the database API. This request is carried out by specifying the time period of interest. In response to the user's request, the database API delivers the data pertinent to the specified period. Leveraging the previously described attributes, users can reassemble the sequence of measurements under investigation and cross-check these data against the $Hash_1$ values. If the $Hash_1$ values demonstrate consistency, users can proceed to extract the $Hash_2$ construction attributes. Upon reassembly of $Hash_2$, users can consult the blockchain API to verify whether these $Hash_2$ values are indeed located within the blockchain. For each $Hash_2$, the API returns a boolean response indicating its presence or absence within the blockchain, thus completing the data immutability verification process.

The strategy of utilizing two distinct hashes fortifies the immutability of the data. Specifically, $Hash_1$ uniquely corresponds to the serialization of data within the one-hour window, while $Hash_2$ incorporates $Hash_1$ as part of its input parameters. This two-tiered hash system effectively enhances the robustness of data integrity verification.

6.2.5 Deployment and Results

The functional evaluation of the proposed architectural framework was carried out in the SSN scenario (refer to 4.5.1). In order to transmit additional data efficiently—data that may be of interest to various stakeholders in a smart city scenario—communication was established through shared access to a 5G network (refer to Fig 6.12).

(a) (b)

Fig. 6.12 Sensor Deployment Test. Air Monitoring Station and third party sensors (a), Final assembly (b).

To generate hashes ($Hash_1$ and $Hash_2$), the SHA256 algorithm [293] was utilized. The hashing process was executed on the intermediary blockchain node. Additionally, a starts system was incorporated to assess the connectivity status and the summarization of MQTT messages containing measurements (refer to Fig 6.13). From this stats HTML window, various parameters can be observed: the existing summarization period (which is 3600 seconds), the total count of MQTT messages received from each sensor, and the number of $Hash_2$ instances that have been entered into the blockchain.

To verify transactions on the blockchain, a blockchain record explorer was used, facilitated by operational dashboards and transaction visualization tools. Utilizing

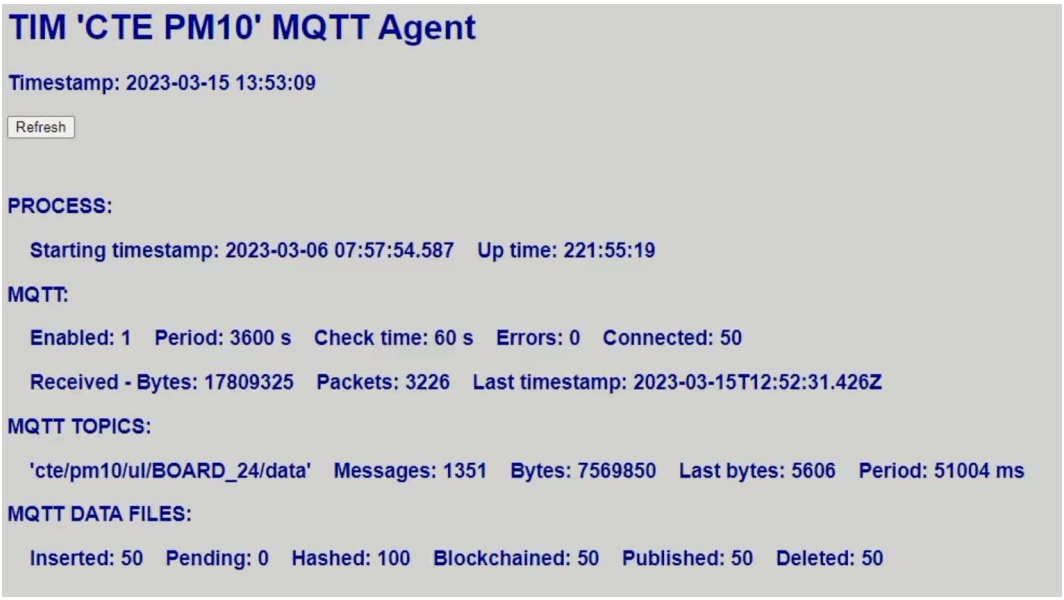


Fig. 6.13 Blockchain Intermediary stats view.

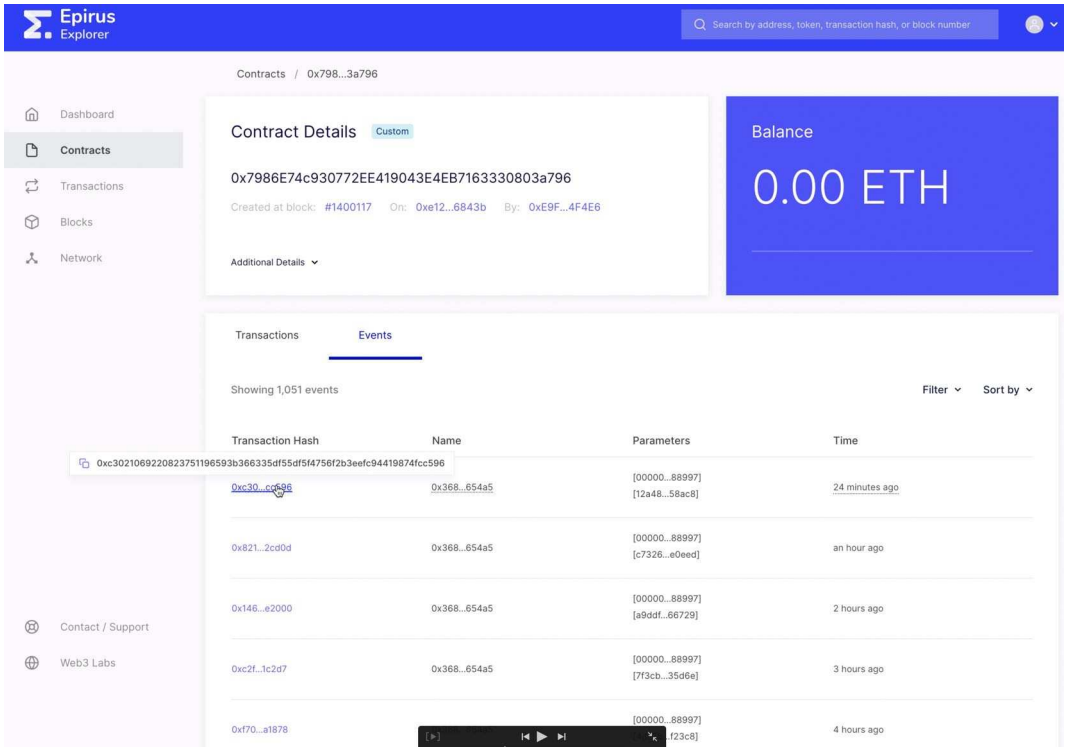
these tools, the hash storage configuration was executed under the scheme of smart contracts (refer to Fig 6.14a). This platform provides the capability to monitor each transaction executed when storing every $Hash_2$ value produced by the blockchain intermediary. Additionally, it facilitates the visualization of the stored $Hash_2$ instances (Fig 6.14b).

Lastly, the querying of the database was accomplished using RESTful requests. These requests were primarily focused on two main resource accessed by the API: *packets* and *packet_summary*. The *packet_summary* resource corresponds to the aggregated packets that the intermediary node generates, while the *packets* resource provides access to the packets transmitted by the IoT devices.

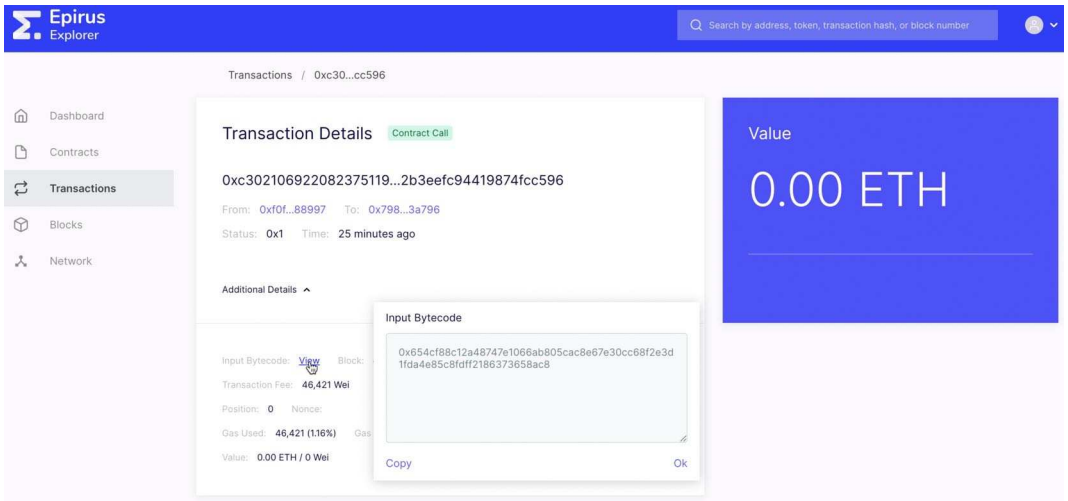
The methods implemented for client access through the database API are detailed subsequently:

GET /packet_summaries/{start}/{end}
Returns the summarized packets generated within the time interval [start, end]. Start and end are Unix timestamps.

GET /packet_summaries/{boardId}/{start}/{end}
Returns the summarized packets generated within the time interval [start, end] from the board identified by *boardId*.
Same format as method *GET /packet_summaries/{start}/{end}*.



(a)



(b)

Fig. 6.14 Blockchain Dashboard. Contract Events Log (a), Contract Transactions Details (b).

GET /packet_summaries/{packetId}
Returns the summarized packet identified by packetId.

GET /packet_summaries/{packetId}/data
Returns the UNENCODED data of the summarized packet packetId as an ordered concatenation of the referenced packet data.

GET /packet_summaries/{packetId}/packets
Returns the packets referenced by the summarized packet in the order defined by the summarized packet.

GET /api/packets/{packetId}
Returns the packet with the specified packetId.

GET /api/packets/{packetId}/data
Returns the data of the packet packetId in the original order in which the packet was created.

During the validation phase, measurements were successfully taken, and the immutability of the data stored on the Ethereum blockchain was confirmed. This procedure enables a user to reconstruct and verify data of interest against the published measurements. As a result, data can be made publicly available, ensuring that it has not been manipulated. This empowers citizens with a reliable source of information on particulate matter density, capitalizing on the benefits of decentralized storage. Furthermore, the utilization of dual hashing enhances the reliability level by ensuring high interdependence between the database and blockchain values, as one relies on the other.

6.3 Conclusions

6.3.1 Blockchain and Smartcities

An analysis of the private BCT architecture was presented in order to achieve power and transaction efficiency in most general government services, where compromising security is not an option. Consequently, the previous requirements were fulfilled by the Tournament Consensus Algorithm (TCA) in a private Blockchain deployment. Subsequently, the TCA was evaluated under a controlled scenario, where its performance for different volumes of transactions was demonstrated. The results

indicate that an adequate amount of computing resources is required by TCA without compromising the security risk associated with the integrity of the chain, in contrast to the analysis conducted on PoW and PoS algorithms, which reveal higher power consumption and security issues in private blockchains.

In accordance with the proposed architecture, the key to initiating a blockchain transaction is held by the smartcards. A robust authentication mechanism was proposed, which combines the three user authentication factors based on the MoC mechanism found in current smartcards and biometric systems. As a result, identity theft for document owners is prevented, and control is maintained to ensure that only the owner can access e-government services.

6.3.2 Air Pollution Monitoring over Blockchain Technologies

The amalgamation of technological trends, such as blockchain and IoT, yields numerous benefits concerning data integrity protection. The proposed architecture provides a proof of concept illustrating the symbiosis between these two technologies. By offering an enhanced level of trust and transparency, this newly proposed architecture adds a robust layer of security, ensuring that the data stored on the system is both reliable and tamper-resistant. This incorporation of blockchain technology acts as a fundamental game-changer, offering a high degree of assurance to all the system users about the authenticity of the data they interact with.

The incorporation of blockchain technology is observed to fundamentally revolutionize the system, providing users with high confidence in the authenticity of the data with which they interact. The presence of metadata within the blockchain allows for the subsequent confirmation of data immutability, which further strengthens the architecture's robustness and reliability. An additional layer of data security is implemented through the dual-storage mechanism deployed in the public blockchain, enhancing the system's resilience to potential attempts of data manipulation by employing the linkage of two distinct hashes, which further complicates data tampering. Moreover, users are provided with the data enabling them to execute their own implementations for calculating each of the different hashes.

While Ethereum's PoS algorithm is more energy-efficient in terms of data integrity, it is imperative to continue evaluating alternatives that could further improve this efficiency. This is crucial to ensure that the drive for data integrity does not

compromise sustainability and efficiency indices. Thus, it is suggested that the architecture presented here may pave the way for future secure, resilient, and reliable IoT systems, setting new standards for data integrity.

Chapter 7

Other IoT Applications

IoT solutions also hold great potential for other dimensions of human development. Referring back to the United Nations Sustainable Development Goals, education is Goal 4, which aims to ensure quality education that is equitable and inclusive. ICT has already made contributions to distance education. However, tasks such as student engagement and obtaining feedback from students are demanding for teachers.

Understanding students' reactions in the classroom is crucial for evaluating the effectiveness of teaching methods. While this task is feasible in smaller class sizes, it becomes increasingly complex when dealing with larger classes, particularly those comprising 50 or more students. To tackle this challenge, this chapter introduces a study that presents a novel IoT system that employs non-invasive techniques, such as facial expression recognition and physiological data analysis, to assist teachers.

The system utilizes a Convolutional Neural Network (CNN) for facial expression recognition and incorporates Photoplethysmography (PPG) to capture physiological data. By leveraging Russell's model, the CNN identifies significant facial expressions, classifying them into active and passive categories. To facilitate comparison and analysis of results from both data sources, operations like thresholding and windowing are applied. When employing a window size of 100 samples, both facial expressions and physiological data indicate an attention level of approximately 55.5% during in-person lecture tests.

By comparing the outcomes of in-person lectures with pre-recorded remote lectures, it becomes apparent that facial expressions, when validated with physiological data, offer valuable insights into students' attention levels within the classroom set-

ting. This research emphasizes the potential of utilizing these combined approaches to enhance teaching effectiveness and gain a deeper understanding of student engagement in educational environments.

7.1 Educational Scenario - IoT System for Affective Learning

Teaching necessitates constant self-evaluation and the ability to independently determine the effectiveness of one's lessons. With experience, educators develop the skills to observe classroom responses to varying teaching methods and course material. While this is relatively straightforward in smaller classes, larger groups of 50 or more students pose challenges to this process, as well as to the teacher-student interaction and the clarity of the lesson's media content.

The scope of this chapter section is to analyse and present an IoT case of use based on the applicability of non-invasive techniques designed to assist teachers in obtaining information about the degree of attention exhibited by their students. Facial expression recognition is tested as primary source, while physiological data analysis is used as validation tool which necessitates the use of a smartwatch, is utilized as a secondary method of validation when it is feasible. To assess the effectiveness of the proposed method, 13 students were involved in two distinct measurement campaigns, which were conducted in different settings: one in-person and the other remote.

Although the methods for conducting facial expression recognition and physiological data analysis have been extensively detailed in existing literature, there are limited proposals regarding the integration of data from both sources to establish a redundant system. We assert that by combining data from these two sources, it is feasible to obtain a complete overview of the degree of attentiveness displayed by students in a classroom.

In relation to facial expression recognition, a Convolutional Neural Network (CNN) was employed to identify eight emotions that are acknowledged across all cultures, such as *happiness, sadness, disgust, anger, contempt, fear, sadness, surprise*, and *neutrality*. On the other hand, PhotoPlethysmoGraphy (PPG) was utilized to measure physiological responses. Subsequently, both methods were processed to

derive a percentage of the level of attention displayed by students during a particular time frame.

7.1.1 Background

The cognitive abilities of humans have been extensively examined from both neurophysiological and psychological perspectives. Various theories and pedagogical methodologies have been developed to enhance the cognitive processes of the human mind [294]. With the increased usage of computers and electronic devices in both in-person and remote teaching settings, there is a concern that these devices may cause distractions that could negatively impact students' level of attention, as evidenced by studies conducted in [295, 296]. Research evidences also suggests that different cognitive processes are strongly associated with physiological responses, including heart rate and eye blinking, among others. By observing these phenomena, it is possible to determine a person's cognitive state, as noted in [297].

The collection and analysis of these factors have been utilized to determine the emotional and attentional aspects that can impact the learning process. In [298], previous learning hypotheses that suggesting a student's attention decreases over time have been refuted through physiological measurements, providing a less subjective means of determining engagement. In a recent review, [299] discusses the various neurophysiological variables used to assess attention levels, which can be categorized into two groups: the central nervous system (CNS) and the autonomic nervous system (ANS). The CNS group is linked directly to the central nervous system and requires invasive techniques for measurement, limiting the comfort of the target student and the scalability of the systems. Examples demonstrating the use of electroencephalography (EEG) to assess attention and learning levels can be found in [300, 301]. In contrast, the autonomic nervous system (ANS) signals, which include facial expressions, eye-based measurements, heart rate (HR), blood volume pressure (BVP), and skin conductance (SC), are less invasive and commonly used to represent and model ANS signals. These variables are useful in detecting attention and emotional factors and can be divided into two main approaches: facial expression detection and multi-neurophysiological signal detection.

Facial Expression Recognition

For many years, techniques for recognizing emotions from facial expressions have been studied. In 1971, Paul Ekman and William Friesen published a list of the six basic emotions that are common to all human groups regardless of their culture: happiness, anger, fear, disgust, sadness, and surprise [302]. These emotions were later expanded by Ekman himself, Daniel Cordaro, and other researchers [303] [304]. Ekman also introduced the Facial Action Coding System (FACS) in 1978, which can be used to categorize the muscular movements of the human face [305]. The FACS system identifies 46 Action Units (AUs) that are responsible for facial movements, and the combination of multiple actions and their respective intensities can generate a vast range of possible facial expressions. However, only experienced annotators can classify facial expressions accurately using FACS.

The authors of [306] outlined the primary channels of emotional communication, identifying facial expressions and vocal tone as the main ones. They also segmented the procedure for recognizing emotions from facial expressions into three steps: detecting the face, extracting facial features, and ultimately describing the emotional state. They stated that the step of describing the affective state from facial expressions is more complex due to its subjective nature. They emphasized that the description of the emotional state should be similar to how a human would describe it. Nonetheless, they pointed out that individual interpretations of emotional expressions can differ depending on factors such as culture and personal experiences. Therefore, they recommended that affect recognition programs should be specifically customized to the context in which they are applied.

Extracting facial features is usually done by identifying AUs from FACS instead of typical expressions of basic emotions like happiness or anger. This is because basic emotions occur rarely, whereas emotions are usually conveyed by variations in specific facial features [307]. Moreover, psychologists study AUs to recognize complicated emotional states [308].

In the paper [309], an attempt was made to address some of the limitations of FACS, such as the absence of detailed spatial and temporal information, by introducing a new representation known as FACS+. Similarly, optical flow-based techniques were employed in [310] to create a time-sensitive approach for recognizing facial expressions.

In a survey conducted by Sariyanidi et al. [308], the authors examine various methods for automatic affect recognition by breaking down their process into four essential components: face registration, representation, dimensionality reduction, and recognition.

In the face registration stage, the face can be registered as a whole or through the combination of different parts or the localization of fiducial points. The next step is to encode the face into a spatial or spatio-temporal representation. Spatial representation encodes image sequences frame by frame and can be performed at various levels of abstraction. Techniques like Local Binary Pattern(LBP), Local Phase Quantization (LPQ), and Gabor Filters are utilized to obtain representations with low abstraction levels. In [311], a combination of Gabor Filters and LBPs is employed for this purpose. Conversely, high-level representation aims to extract semantically interpretable features. Regarding dimensionality reduction, it can be achieved through either pooling, feature selection, or feature reduction. Finally, in the recognition phase, the result is produced as labels indicating emotions or facial actions, including AUs. Techniques such as Hidden Markov Model (HMM), Dynamic Bayesian Network (DBN), Support Vector Machine (SVM), Relevance Vector Machine (RVM), and Conditional Random Field (CRF) are commonly utilized for this purpose.

Another aspect worth noting is the range of datasets employed for training and validating affect recognition methods. One key differentiation lies in the labeling of the dataset, with some utilizing FACS [312], whereas others report prototypical expressions [313, 314]. Certain datasets comprise of posed expressions [312, 313], occasionally enacted by actors [314], while others are created via labeling natural images. Additionally, these datasets may comprise of subtle expressions [314] or exaggerated expressions [315, 316]. Images can be specifically generated for emotion recognition purposes or collected from the internet and movies [317].

The use of neural networks has become more prevalent in recent years, and many researchers have explored their potential in enhancing affect recognition tasks. One such study, conducted by Liu et al. [311], combined traditional techniques such as Gabor Filters and LBPs with Extreme Learning Machines (ELM) to create a more standard pipeline that is suitable for real-time applications. The usage of Deep Learning models for affect recognition is limited by the scarcity of available images, as stated in [318]. To address this issue, a Shallow CNN is employed in

[319] to recognize micro-expressions with a limited number of training samples. In contrast, [318] tackles this problem by utilizing domain knowledge to strongly guide the model towards the relevant facial areas. A study by Liu et al.[320] proposed a Deep Belief Network (DBN) that unified the three main stages of affect recognition, namely feature learning, feature selection, and classifier construction. On the other hand, Meng et al. [321] introduced an identity-aware CNN to address variations caused by personal attributes and achieved better performance in facial expression recognition. In another work by Sini et al. [322], an Artificial Neural Network (ANN) was trained using an ensemble of databases to reduce the model's sensitivity to image variability and increase the number of available samples. Rifai et al. [323] used a Contractive Convolutional Network (CCNET) to obtain translation invariance of the facial features in the image.

Multi-Neurophysiological Signal Recognition

Another method of researching affect recognition involves measuring various ANS neurophysiological signals, which provide a reliable measure of emotion and attention.

In the paper [324], the authors proposed a cyber-physical social system that utilizes multiple sensors and cameras along with a quiz creator to monitor the learning progress of students. They employed reinforcement learning techniques on the data, which enables the teacher to increase student engagement. This algorithm utilizes the students' heartbeats, eye blinks, and facial expressions to suggest the next learning activity depending on their current learning state. In the study presented in [325], the authors proposed an automated system to classify student engagement while they performed a writing task. To achieve this, computer vision techniques were employed to extract three different sets of features from video recordings: heart rate (HR), facial features, and Local Binary Patterns in Three Orthogonal Planes (LBP-TOP). Engagement-detection models were created for each of these features, as well as two fusion models that used all the features or only a subset of them. However, the method faced limitations in accurately extracting facial features and estimating heart rate from video recordings. The authors of [326] presented a model for e-learning that takes into account the mental and emotional state of the learner and provides learning activities accordingly. This model utilizes biophysical variables,

such as HR, SC, and BVP, to automatically select lessons based on the learner's affective state.

A rising methodology is the use PPG technology. It is an electro-optical technique that is non-invasive and can measure changes in blood volume in the microvascular tissue beneath the skin. This method uses light to detect the pulsatile nature of the circulatory system and measure the beats associated with changes in the bloodstream's bulk. It is considered a low-cost and simple technique [327].

The PPG waveform's pulsatile component is the Alternating Current (AC), and its fundamental frequency typically relies on the heartbeat frequency, which is generally around 1 Hz. After proper electronic filtering and amplification, both AC and Direct Current (DC) components can be extracted for further analysis of the pulse wave. PPG technology requires minimal optoelectronic components, namely a light source to illuminate the tissue, such as tissue or skin, and a photodetector to measure the small changes in light intensity caused by the blood volume changes in the tissue. Several factors can influence the reproducibility of PPG measurements, such as the method used to attach the probe to the tissue, the pressure applied at the probe-tissue interface, the pulse amplifier and bandwidth, the reduction of movement artifacts, the subject's posture and relaxation, breathing, wakefulness, room temperature, and acclimatization. The use of LED, photodiodes, and phototransistors in semiconductor technology has improved the design of PPG probes, resulting in smaller, more sensitive, portable, reliable, and reproducible devices. PPG technology is extensively employed in clinical settings for monitoring essential physiological parameters such as blood oxygen saturation, blood pressure, heart rate, and heart rate variability [328]. One of the primary advantages of PPG technology is that it enables the measurement of these parameters without requiring electrodes or complex movable equipment, making it suitable for use in primary care and community-based clinical settings. Additionally, PPG technology uses simple, low-cost, and compact semiconductor devices, which make it a cost-effective option. Finally, PPG technology is an advanced computer-based pulse wave analysis technique that can provide accurate measurements of physiological parameters.

In clinical settings, Heart Rate (HR) is an essential physiological parameter that can be measured by analyzing the AC component of the PPG pulse, which is synchronized with the heartbeat. To improve the accuracy of HR detection, digital filtering and zero-crossing techniques can be applied to separate the HR and

respiratory components from the PPG signal. Additionally, Blood Pressure (BP) has been extensively studied in vascular disease research. An algorithm based on pulse arrival time can be used to calculate BP and compared with traditional methods.

The regulation of heart rate (HR) and blood pressure (BP) is crucial for maintaining BP homeostasis, and BaroReflex Sensitivity (BRS) is often used to characterize this regulation in milliseconds per millimeter of mercury ($ms/mmHg$). Its non-invasive assessment is possible using finger-pressure cuff/PPG technology to measure beat-to-beat peripheral arterial BP waveforms [328]. In a study by Drinnan et al. [329], it was demonstrated that paced breathing at a rate of 0.1 Hz distinctly modulated the Pulse Transit Time (PTT) at the finger and ECG RR interval. Using an overlapping waveform morphology technique, the incremental PTT and beat-to-beat interval (RR) values of ECG on a beat-to-beat basis, and their variability were summarised using simple statistics. The cardiovascular system's non-stationary balance fluctuates between consecutive heartbeats and HRV intervals.

The HRV is a collection of measures utilized to estimate the activity of the ANS. It can be represented as features of the PPG power spectrum, such as low frequency (LF) and high frequency (HF), or as properties of the Beat-to-Beat (BB) interval. The BB features are commonly used in the field of emotion research, and some of them are defined as follows [330]:

- SDNN - Standard Deviation of BB intervals.
- SDSD - Standard Deviation of Successive BB interval Differences.
- RMSSD - Root Mean Square of Successive BB interval Differences.
- pNN50 - Percentage of adjacent BB intervals that differ from each other by more than 50 ms.

Currently, there is a growing interest in academia and the market regarding the design and development of wearable biosensors, owing to their potential use in human health monitoring and personalized medicine. Generally, Wearable BioSensors (WBSs) are portable electronic devices that integrate sensors into or with the human body in various forms and shapes such as watches, gloves, clothing, and implants [331]. These smart portable devices analyze and record live sensing data of human biological parameters, such as blood pressure, heart rate, and temperature, which

are of significant value in healthcare applications [332]. Furthermore, to obtain a reliable and high-quality signal, traditional PPG sensors require a firm attachment to the skin; wearable sensors with measuring functions have helped to solve and identify various significant problems in the healthcare, medical, and sports fields. In the study presented in this study, smartwatches, commonly known as wrist-worn wearable devices, were utilized to record heart rate signals.

Internet of Things for Affective Learning

In the current age of the Internet of Things [333], also known as the Internet of Everything, countless devices are linked to the internet and performing specific functions. These devices are utilized for personal use or for the purposes of companies or communities, such as monitoring fitness or the environment. The data generated by these devices can be utilized for making decisions or adjustments to one's lifestyle. However, the extensive range of sensing and actuating devices available face scalability and system integration challenges. Therefore, concepts such as the Social Internet of Things and corresponding regulations have been introduced to address these limitations. [334–337]

The integration of IoT technologies in the field of education is becoming increasingly popular as a means to enhance traditional classroom settings. This technology enables the collection and analysis of data from educational contexts, ultimately aiding in the facilitation and customization of learning processes [338]. The objective of this study is to harness the potential of IoT to optimize the collection of feedback information for teachers seeking to maximize student engagement. Other researchers have already explored the merging of IoT and education domains. For example, in a study conducted by Tan et al. [339], they employed a WiFi-enabled Radio Frequency Identification (RFID) reader to assist teachers in automatically recording attendance to lessons and monitoring students' behavior. In addition, they integrated Quick Response (QR) codes to enable students to quickly access course materials and provide real-time interactive responses to stimulate participation. The implementation of this system reportedly resulted in an attendance rate increase from 85% to 98%. The authors of a different article [340] created a Google Home voice assistant application that can communicate with teachers and students based on the context of the textbook. The model utilizes Machine Learning techniques to

recognize user sentences and incorporate new ones. The study consisted of two case studies conducted in Bangladesh, and the outcomes were presented.

Specifically, to improve affective learning, Yadav et al. [341] proposed an IoT-based framework that utilizes wearable devices to detect human emotions, including heart rate, movement, and environmental data such as temperature and humidity. The data gathered is then used to monitor students with attention deficit hyperactivity disorder and provide feedback to educators. Awais et al. [342] proposed an alternative IoT framework that uses datasets from diverse ANS and CNS physiological signals to develop an emotion recognition system. The system analyzes the data and transmits it to an IoT hub for real-time communication to identify emotions. The authors aim to provide a solution for remote learning and healthcare support during the COVID-19 pandemic.

7.1.2 Methodology

The methodology suggested follows a redundant system approach, utilizing two sources of data. In order to enhance the comprehensibility of the study, this subsection has been subdivided into three principal components:

1. Recognition of facial expressions using neural networks.
2. Physiological data analysis.
3. Merging of the information.

Recognition of Facial Expressions Using Neural Networks.

As discussed in subsection 7.1.1, recognizing facial expressions can be achieved using artificial intelligence (AI) systems. With a well-designed training phase, these systems can learn to identify people's facial expressions in images or videos. Neural networks, which offer high precision, are one of the techniques commonly used in AI. The method proposed in this study is based on the neural network developed and trained in [322]. Interested readers can refer to this paper for additional details.

To train a machine learning system to differentiate between a limited number of classes, labeled examples are necessary. For instance, a dining chair, rocking

chair, and office chair are all chairs despite variations in their size, shape, and color. Similarly, a dining table, living room table, and office table can also be categorized as tables. It is crucial to use categorization supergroups and visually represent the various characteristics an object may possess to aid the machine learning system. Consequently, the labeling process must be meticulous to minimize errors to the extent possible by humans [343].

Training a Neural Network Through Quality Image Database Over the years, numerous databases containing high-quality facial expressions have been made available. In the study by Sini et al.[322], various neural networks based on the CNN model proposed by Ferreira et al. [318] were trained using the Keras library [344]. The authors used a combination of different facial expression databases, which they referred to as Databases Ensembles (DEs), in the training process. Using multiple databases can enhance the performance of the trained system by:

- have a different photographic quality;
- have different backgrounds;
- portray multiple human traits;
- increase the size of the train, validation, and test dataset.

This enables the development of AI systems that can generalize better and maintain reliability even in situations that differ greatly from the training conditions. As described in the subsection 7.1.1, facial expressions can be either spontaneous or posed. Therefore, two distinct DEs were generated:

- *Ensemble 1*, composed solely of image databases that feature individuals displaying both posed and spontaneous facial expressions.:
  - Extended Cohn-Kanade Database (CK+) [312] [345];
  - FACES Database [346];
  - Facial Expression Recognition 2013 Database (FER2013) [315] plus FER+ annotations [316];
  - Japanese Female Facial Expression (JAFFE) [313];

  - Multimedia Understanding Group Database (MUG) [347];
  - Radboud Faces Database (RaFD) [348].
  - Static Facial Expressions in the Wild (SFEW 2.0) [317].

- *Ensemble 2*, composed of image databases that solely depict people with posed facial expressions:
  - CK+;
  - FACES Database;
  - JAFFE;
  - MUG;
  - RaFD.

The *Ensemble 1* DE is made up of 7 distinct databases, totaling 43993 images, with 8 different facial expressions representing emotions: *happiness, sadness, disgust, anger, contempt, fear, sadness, surprise*, and *neutrality*. The creation of the *Ensemble 1* was facilitated by the use of the Facial Expression Database Classifier (FEDC) [349], a software that aids in the creation of DEs. The images in the DE were preprocessed by following the steps listed below:

- Grayscale conversion;
- Resize to 48x48 pixels;
- Face cropping;
- Z-score normalization;
- 90%-10% random split for train and test datasets.

To enhance the precision of the system and acquire the validation dataset, a 9-fold cross-validation technique was applied to the initial subset, which resulted in a final division of 80%-10%-10%. The following data augmentation techniques were used during each training epoch:

- Random horizontal flip;
- Brightness range among 50% and 100%;

- Shear range among ±2.5%;
- Random rotation among ±2.5 degrees;
- Zoom transformation interval among ±2.5%;
- Width and height shift range among 2.5%.

The CNN described in [318] achieved a test accuracy of 80.38% with the best fold combination, using *Ensemble 1* as the train, validation, and test dataset along with the settings mentioned earlier. The confusion matrix for this result is shown in Figure 7.1.

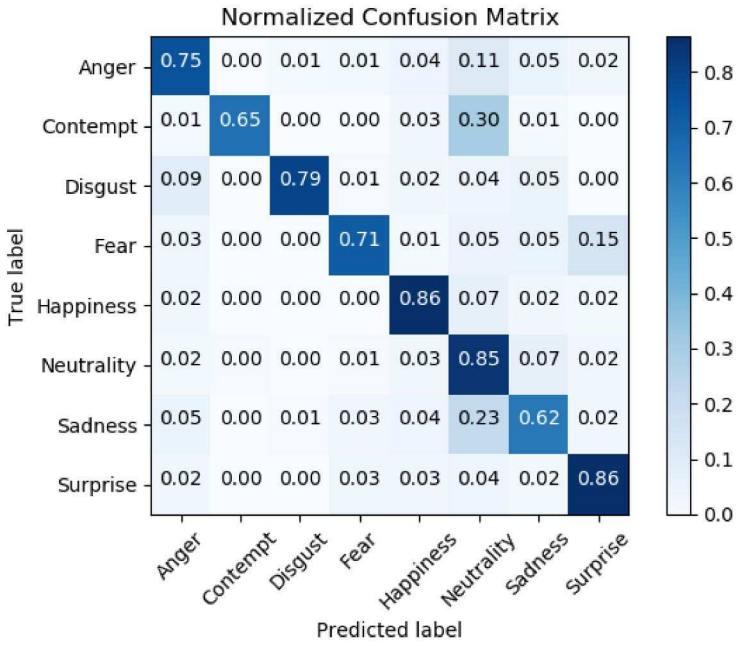


Fig. 7.1 Normalized confusion matrix of the neural network described in [318], trained by using the *Ensemble 1* dataset [322].

The achieved accuracy is relatively low when compared to the accuracy obtained from using only CK+ dataset, which was 92.5%, and *Ensemble 2* dataset, which was 96.78%. However, the authors have identified some issues with these datasets: The first issue is that they have a limited number of images, which reduces the neural network's ability to generalize and perform well in real-world scenarios. The second issue is that they were captured in a studio setting, which results in less variation than in real-world scenarios. Finally, only posed facial expressions are included in

these datasets, creating a bias towards spontaneous expressions that are more natural but have less inter-class variability. [322].

Despite having lower quantitative metrics compared to the other datasets, the neural network trained using *Ensemble 1* as the train, validation, and test dataset was deemed reliable for the facial expression recognition problem at hand. Hence, the neural network trained on *Ensemble 1* was selected for use in the proposed approach.

Facial Expressions Prediction and Post-processing The process of predicting facial expressions was carried out using Emotion Detector (ED) software [350] [351], which was created using: *Java, Apache Maven, OpenCV* and *DeepLearning4J (dl4j)* [352]. It is capable of detecting facial expressions of an individual through input from a camera, video, or image. When a video is given as input, a sampling rate must be chosen since the neural network was trained to analyze single frames. However, regardless of the input type, the software is able to:

1. Displays the current image or frame;
2. Performs face detection using Haar Cascades [353];
3. Updates the interface by drawing the detected face in an ad hoc box. Otherwise, *Face not found* warning is shown;
4. Performs facial expression recognition using the neural network discussed in subsection 7.1.1;
5. Updates the graph by showing the newly obtained result;
6. Optionally, it saves the result as a screenshot or in a Comma-Separated Values (CSV) format file.

To enhance ED for this project, it was decided to incorporate the option to choose between using Haar Cascade Classifiers or a pre-trained Deep Neural Network (DNN) for face detection, with the latter offering greater precision. The DNN was an Single Shot multibox Detector (SSD)[354] based on ResNet-10 architecture [355], trained with the Caffe framework [356] to detect faces on 300×300 images. Facial expression analysis utilized the DNN setting and analyzed videos from both

in-person and remote lessons by capturing two samples per second. The resulting data was exported to CSV format for further analysis.

To align the predictions of facial expressions obtained through the CNN with physiological data, they were categorized into two moods, which followed a classification similar to Russell's [357]:

- Active:
  - Anger;
  - Fear;
  - Happiness;
  - Surprise;
- Passive:
  - Contempt;
  - Disgust;
  - Sadness.

The first group is associated with high arousal moods, referred to as active, while the second group is associated with low arousal moods, referred to as passive. The study did not take into account the pleasure-displeasure characteristics.

Facial expressions that show neutrality have been left out of this categorization since they cannot be classified as either active or passive. Next, the level of activation was calculated by tallying the number of active mood readings in a given time period (in this case, 100 seconds) and scaling it to $Ws_{NN} = 300$ (meaning that 300 indicates an interval with only active moods, while 0 indicates an interval with only passive moods).

A binary classification was applied to the activation level obtained from facial expressions. This was done by comparing the activation level with a threshold value (T_{fe}), which was set at $Ws_{NN}/3$. The resulting binary array, called the Attention Array from facial expressions (AA_{fe}), contained a value for each time interval indicating the number of active moods in the last 100 samples. If the grade of attention was higher than the threshold, the window was labeled as *Attention*, otherwise it was labeled as *Distraction*.

Physiological Data Analysis

The Attention Detection algorithm that relies on physiological reaction data is explained in this segment. The data were gathered using a group of consumer grade smartwatches (Garmin Venu Sq) that transmit data through BLE (Bluetooth Low Energy) to a group of smartphones, with a sampling frequency of 1 Hz. The recorded data is saved in a text file. The datasets that were collected consist of both heart rate (HR) and heart rate variability (HRV). MATLAB was used to develop the data collection algorithm. Prior to executing the algorithm, a two-level calibration phase is performed on each patient to estimate the thresholds for emotional stage detection:

1. Based on their information, which are age, gender, weight, and height;
2. From their initial condition estimation.

The algorithm is capable of performing real-time behavioral analysis during lectures by monitoring the HR and HRV for a specified number of samples (N), referred to as the Window Size (WS). From this observation window, only specific emotional stages, namely SDNN, RMSS, and SDSD, are used. These stages were described in 7.1.1. By sliding the window with a particular initial delay, an output can be generated every second.

The Grade of Attention (GA) can be increased by comparing the instantaneous value of the three parameters with the values obtained during the calibration phase. To compare the GA values, a threshold (T_p) is calculated, which is determined as follows:

$$T_p = \frac{3}{2} \cdot WS$$

where 3 refers to the number of the calculated emotional stages. Thus, the algorithm generate two levels of attention stored in an Attention Array based on physiological reactions (AA_p). If the grade of attention is higher than T_p, the window is labeled as *Attention*, otherwise as *Distraction*, similar to the activation levels obtained from neural network classification.

Merging Physiological and Facial Expressions Data

The process of obtaining AA_{fe} and AA_p is necessary to enable the comparison of data from disparate sources and to apply the windowing algorithm. The study's objective

is not to examine the attention of individual students but to gather aggregated outcomes. There are several reasons for not examining a single student, which are as follows:

- The system outcomes are not influenced by the engagement level of an individual student towards the lecture topic.
- A student's interest in the lecture may vary on different days due to personal factors such as drowsiness and mood.
- Somatic features that vary significantly from the training data can impact the performance of the CNN used for facial expression recognition.

An essential factor to determine is the window size (WS). The windowing algorithm produces an output indicating the number of *Attention* labels within the sliding window. This window extends to the current time and begins WS samples earlier. Since people's facial expressions change rapidly, the neural network output may exhibit rapid fluctuations. To smooth out the results, a 5-minute attention counter filter is applied to obtain the averaged activity level, known as Attention Behavior (AB). Consequently, the algorithm will begin functioning effectively after 5 minutes. To maintain consistency, the same filtering operation is also applied to the physiological data. Figure 7.2 provides a sequential overview of the proposed approach.

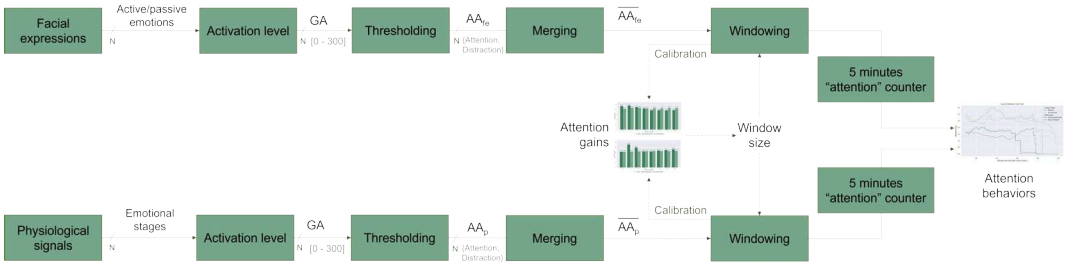


Fig. 7.2 The workflow of the proposed analysis approach. The dashed arrows represent the WS calibration feedback.

Comparison of Methodology with Other Solutions

The objective of this project is to utilize physiological data to validate facial expressions, with the ultimate goal of developing a system that relies solely on facial

expressions. The advantages of implementing such a system are: Firstly, it offers a less invasive approach compared to other methods, ensuring user comfort and minimizing any potential discomfort. Additionally, this system can be applied in various scenarios, whether it be in-person lectures or remote learning environments, providing versatility and adaptability. Furthermore, its cost-effectiveness is worth noting, as it primarily relies on readily available hardware such as webcams or similar devices, reducing the need for expensive equipment or infrastructure.

These factors are noteworthy as the aforementioned systems relied on a complex multimodal acquisition methodology that is not readily feasible for everyday use. In contrast, this project utilized consumer-grade devices, making it more cost-effective. Furthermore, advancements in technology enabled comparable results to be achieved with fewer parameters. Specifically, while previous work such as [326] utilized measures such as BVP, SC, and HR, this project focused solely on HRV and HR for its analysis.

An additional distinction is the ability of the facial expression recognition system to effectively handle multiple individuals. This is achieved through the utilization of two separate CNNs: one dedicated to extracting facial features and the other focused on recognizing facial expressions. As a result, this system can be utilized without any difficulties during both in-person lectures and remote sessions, where multiple individuals are present and engaged in the learning experience.

Another instance can be observed in the study conducted by Monkaresi et al. [325], where tests were carried out with a comparable number of students but in a distinct context involving a writing activity. In this particular study, facial tracking was accomplished using a Kinect device, while HRV and HR data were obtained through the use of ECG. However, the challenge with these technologies lies in their limited applicability within a realistic environment. Additionally, engagement levels were estimated using LPB-TOP and a machine learning tool employing customized classifiers.

7.1.3 Experimental Results

According to the background subsection, previous studies [295, 296] have indicated that the level of attention in online classrooms is typically lower compared to in-person classes. This is attributed to distractions caused by electronic devices,

including the devices used to participate in the lecture. Conversely, when students and professors are physically present in the same room, there is generally a higher level of interaction [358].

Considering these factors, the decision was made to assess the level of attention in both lecture modes. A group of 13 participants, consisting of 4 females and 9 males aged between 21 and 35, was enrolled for the study. The first phase of the study involved in-person lectures, during which the following data were collected:

- Physiological data, using commercial smartwatches;
- Facial expressions, using visible-wavelength cameras.

The second campaign involved remote lectures, and the following data were collected:

- Physiological data, using commercial smartwatches;
- Facial expressions, using visible-wavelength cameras.
- Reaction times, using a dedicated application for this specific task.

In-Person Lectures

Two 1-hour in-person lectures were conducted, with each lecture having 4 participants, resulting in a total of 8 volunteers. The lectures were facilitated by two experimenters, where the first experimenter acted as the teacher, and the second experimenter programmed the teacher's wristwatch to provide discreet vibrations at specific times that were unknown to the teacher.

When the alarm went off, the teacher instructed the volunteers to perform a simple action, such as touching their shoulder, nose, or head, and asked if they were paying attention to the explanation. The reaction times were recorded by analyzing the video footage of the lecture, starting from the moment the teacher made the request until the volunteers completed the actions. To ensure a fair comparison, only the participants in the first experiment's group were informed about the request.

Remote Lectures

Remote lectures, which gained popularity following the outbreak of the SARS-CoV-2 pandemic in 2020. To experiment in this scenario, 6 volunteers participated in a pre-recorded lecture lasting approximately 50 minutes. They used their personal devices to record their facial expressions and were equipped with smartwatches to monitor their physiological reactions. To compensate for the absence of a lecturer, an application was employed to randomly inquire whether the students were actively engaged in the lecture or not.

Application to Measure the Response Times

The Reaction Time Tool, an application utilized in the study, was created using the *Microsoft.NET 6* framework and implemented using the *C#* programming language. The application's Human-Machine Interface (HMI) was designed within the Windows Presentation Foundation (WPF) environment.

The HMI consists of two primary windows. The first window appears upon launching the program and prompts the volunteers to specify the location for saving the log file. It also features a blinking hourglass, serving as a visual indication that the application is functioning properly.

At predetermined but undisclosed intervals, a message window appears to the volunteers, inquiring about their level of attentiveness during the lecture. The user has the option to respond with either "Yes" or "No" based on their attention level. The window cannot be closed without providing a response. To capture the user's attention when the message window appears, the application emits a distinct sound. This sound has been deliberately chosen to be distinguishable from the typical sounds produced by standard operating system message windows.

Each time the window is displayed, the application records a timestamp of the event in the log file. Another event is recorded when the window is closed, along with the corresponding response from the volunteers. Additionally, for ease of data analysis, the application calculates the time difference (Δ time) between the display of the window and the user's response.

Moreover, in this experiment, one group consisting of 4 individuals is informed in advance about the presence of these message windows, whereas the other group of 2 participants is not aware of them.

It is anticipated that, starting from the second occurrence, the users will recognize these windows more quickly compared to their initial encounter.

To summarize, the application serves two primary objectives:

1. It prompts the students to indicate their level of attentiveness during the lecture.
2. It logs and measures the reaction time of the students.

The source code for this application can be accessed by interested readers on GitHub under the MIT license [359].

Self-Assessment Results

To compare the attention levels derived from both physiological data and facial expressions in both the in-person and remote lectures with the self-assessment provided by the volunteers, the following results were obtained.

In the case of the in-person lecture, when the students were asked if they were paying attention, it was observed that 5 of them consistently answered yes, while 3 always responded with no. These affirmative answers accounted for approximately 40% of the total responses.

During the remote lectures, each volunteer was asked the same question 5 times using the Reaction Time Tool. The results showed that there were 20 affirmative answers and 5 negative answers. However, it was noticed that 2 out of the 20 affirmative answers had unusually long reaction times, indicating a lack of attention. Therefore, these two answers were considered false. Taking this into account, there were 18 affirmative answers and 7 negative answers, resulting in an attention level of 72% when using a window size of 10 minutes.

This finding deviates somewhat from the measurements obtained with the proposed approach, which indicated an attention level of approximately 52%. However, it is important to note that the small number of volunteers limits the statistical significance of this comparison.

Window Size Determination

The initial challenge involved determining the appropriate window size (WS) for the windowing algorithm. Since there is limited information in the literature regarding behavioral analysis of emotional stage parameters, the WS was determined empirically. The value was chosen based on minimizing the variations observed when applying the windowing algorithm to $\overline{AA_{fe}}$ and $\overline{AA_p}$ readings.

Figure 7.3 illustrates the correlation between the two systems. The Attention Gain (AG) represents the percentage of *Attention* states detected from the raw Attention Behaviors (ABs), without applying the 5-minute attention counter. The graphs demonstrate that the optimal results for in-person lectures were achieved with a WS of 100 samples. Therefore, the AGs from the in-person lectures were selected as they yielded better performance with the Neural Network. In remote lectures, there is a notable prevalence of *Neutrality* outcomes, indicating that the results from this source are less predictive compared to the results from Physiological Data.

By considering 100 samples from both sources, the Root Mean Square Error (RMSE) can be computed using the following formula:

$$RMSE = \sqrt{\frac{(\hat{AG} - AG_p)^2 + (\hat{AG} - AG_{fe})^2}{2}}$$

With AG_p representing the Attention Gain from physiological data, AG_{fe} representing the Attention Gain from facial expressions, and $\hat{AG}$ denoting the average of AG_p and AG_{fe}, the Root Mean Square Error (RMSE) values were calculated. The RMSE for in-person lectures was found to be 0.02, while the RMSE for remote lectures was determined to be 4.76.

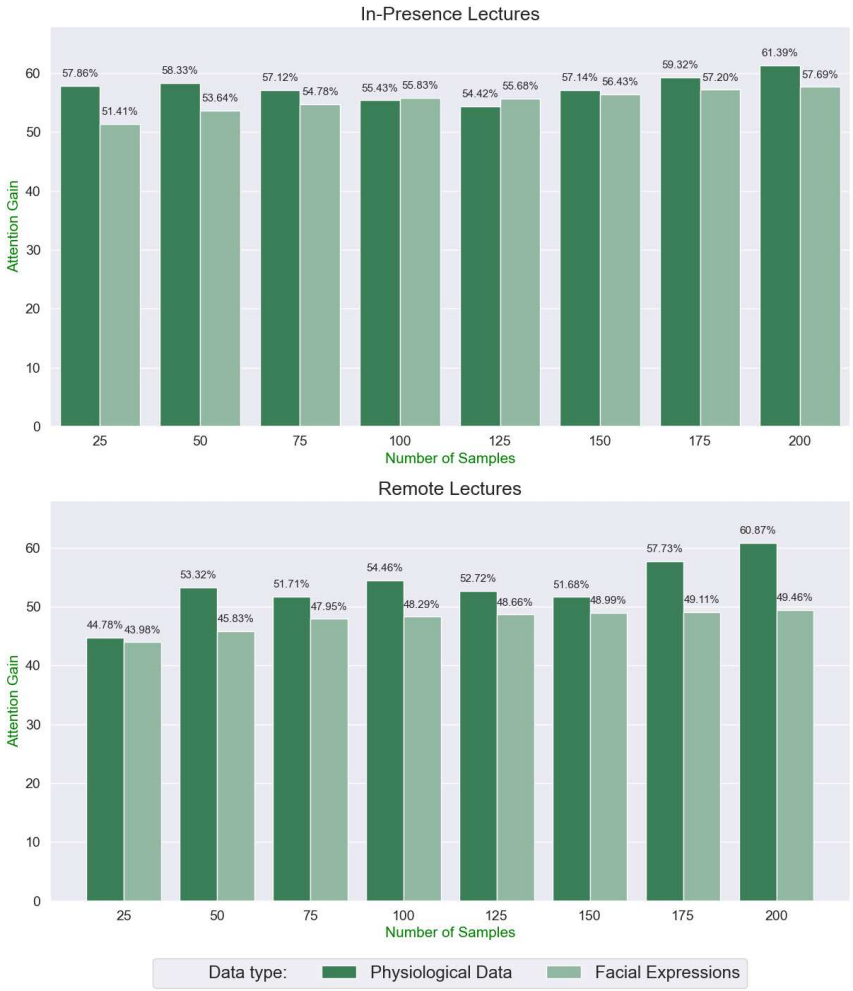


Fig. 7.3 Comparison of the AGs obtained from the $\overline{AA_p}$ and $\overline{AA_{fe}}$ during the in-presence and remote lectures with different windows sizes.

Attention Behavior Analysis

After determining the WS, the AB can be computed. T To analyze the collective AB of the students, the results of the windowing algorithm need to be filtered. The results presented in Figure 7.4 were obtained by counting the occurrences of *Attention* states detected in the preceding 5 minutes of AA_i.

Although reaction times were collected, they were not utilized in this study as there is no existing literature establishing a correlation between attention level and reaction times. However, the data collection of reaction times is included in this paper to describe the perturbation caused by attention checks on student attention

levels, which may be significant for replicating these experiments. As mentioned earlier, the data became available only after 5 minutes from the beginning of the lecture. However, this delay does not impact the analysis of AB, and considering that lectures typically extend beyond 1 hour, it is deemed acceptable for practical implementation in real-world scenarios.

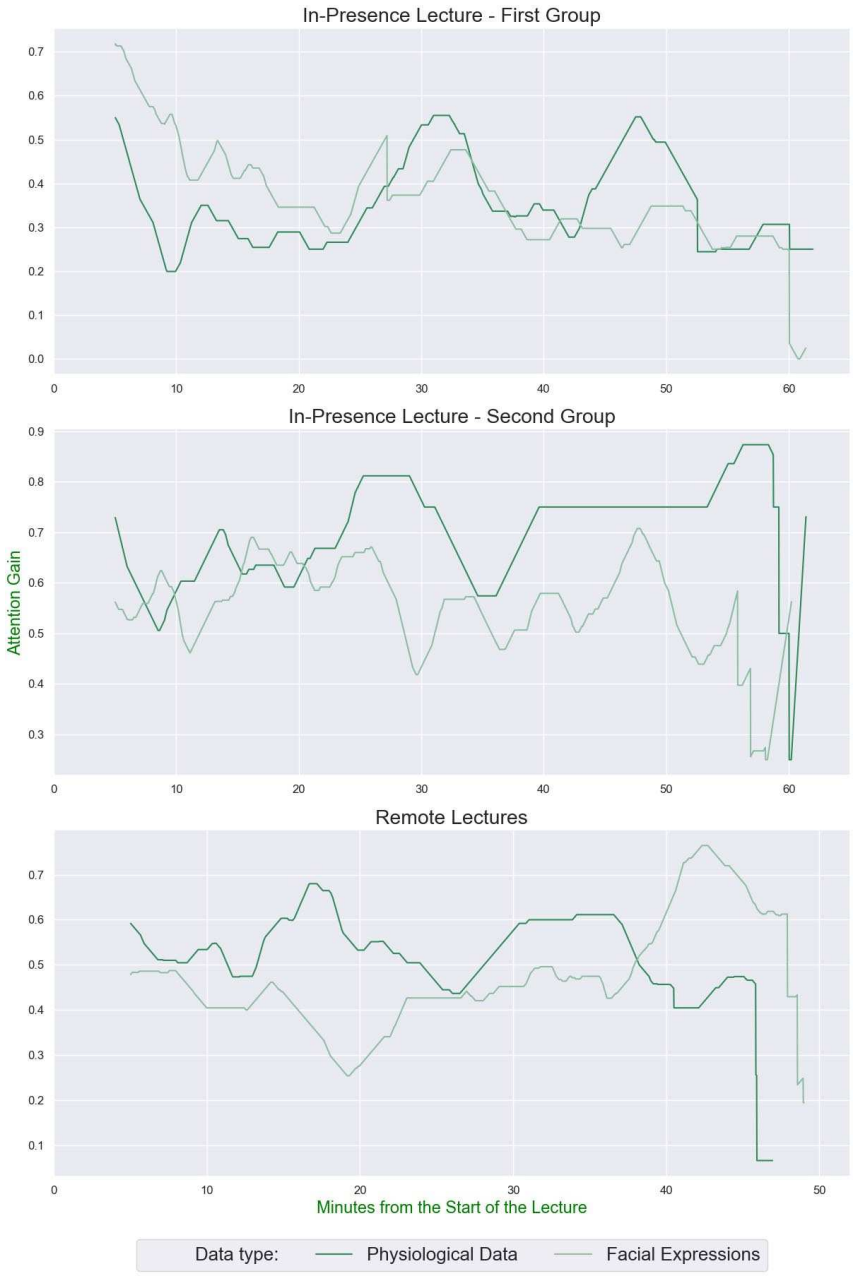


Fig. 7.4 ABs of the in-presence and remote lectures obtained from $\overline{AA_p}$ and $\overline{AA_{fe}}$ with a WS of 100 samples and applying the 5 min attention counters every second.

Results Discussion

The findings of this study offer valuable insights into the attention behaviors of students in both in-person and remote lectures. As discussed in the subsection 7.1.3, it is evident that there are distinct differences in the levels of attention observed between the two types of lectures. Furthermore, the calculated ABs from both data sources align with each other, providing further support for the consistency of the results.

Figure 7.4 illustrates the average attention levels of the in-person and remote groups. From the graph, it is evident that there exists an attention disparity between the two groups. The average values suggest that the second group exhibited a higher level of attention compared to the first group. This observation can be attributed to several factors. Firstly, the first group was aware that actions may be requested during the lecture, which could have conditioned their state of attention. Secondly, it is possible that the second group had a greater interest or engagement with the lecture topic compared to the first group.

In remote lectures, the prevalence of neutral facial expressions can be attributed to the passive nature of the lessons. As a result, the use of Physiological Data becomes particularly valuable in assessing the attention levels. Regarding the in-person lectures, there is an interesting pattern observed in the peak of attention. The second group reaches its peak after approximately 20 minutes, while the first group reaches it after around 30 minutes. This difference in timing suggests that the second group found the lecture topic more captivating or engaging compared to the overall average.

Additionally, during one of the in-person lectures, there was a high level of interaction between a student and the teacher. In the other lecture, after 25 minutes, there was a request for the students to touch their heads, which may have caused a momentary shift in attention.

The consistent results observed in the remote lectures can be explained by the technical nature of the topic (i.e., designing an IT component), which typically results in less variation in engagement levels compared to the in-person lecture, which focused on informative content (i.e., how airplanes can fly).

Lastly, it is noteworthy that the results obtained in this study align with those reported in the two works discussed in the last part of subsection 7.1.2.

7.1.4 Conclusions

The research presented in this chapter specifically addresses the redundancy of two distinct systems that can be utilized when ground truth is unavailable. The primary objective is to measure the level of attention among students in two different scenarios: in-person lectures and remote lectures.

The study introduces two key technical contributions. The first contribution involves the utilization of a Convolutional Neural Network (CNN) for facial expression recognition, while the second contribution entails the development of a deterministic algorithm designed to analyze physiological reactions.

These two methodologies, which are widely recognized and accepted, have been integrated in a novel and redundant approach to merge data from facial expression recognition and physiological analysis. The CNN is responsible for classifying emotions, while the physiological algorithm determines whether the individual is in a state of *Attention* or *Distraction*. Significant efforts were dedicated to generating data in the form of attention arrays, which facilitate the comparison and analysis of the combined data from both sources.

The proposed IoT system has the potential to assist teachers in assessing the level of attention in their classrooms. By providing attention feedback, it can identify moments during lessons that require attention and help teachers improve the quality of their teaching or enhance student engagement.

It is essential to note that the system does not generate data specific to individual students' attention levels. Therefore, it is crucial to avoid sharing data obtained prior to the merging process. This data aggregation approach also addresses privacy concerns, ensuring the confidentiality of individual student information.

As described in the previous subsection, the experimental campaign revealed a significant occurrence of Neutrality outcomes. This circumstance diminishes the reliability of the results obtained through the Neural Network when compared to the analysis of Physiological Data. Enhancing the performance of the Neural Network could address this concern and improve its accuracy. In the pursuit of redundancy, additional data sources can be incorporated. For instance, the inclusion of a blink counter and an eye gaze direction controller could be explored as potential extensions to the system. These additions would provide further insights into students' attention levels and enhance the overall analysis.

Chapter 8

Conclusions and Outlook

The recent pandemic and the ongoing climate emergency have demonstrated the urgent need for a significant transformation of urban environments. The goal is to promote Healthy Cities aligned with the Sustainable Development Goals. Interestingly, the pandemic has had certain positive impacts, such as reduced car traffic leading to improved air quality and perceived road safety, thus enhancing the overall quality of life. Moreover, the resulting improvements in air quality and the promotion of pro-environmental citizen behaviors may generate positive externalities, including potential economic benefits from reduced healthcare costs.

Rapid advancements in technology have facilitated a better understanding of the interconnected phenomena within cities. Here, ICT and IoT play crucial roles in enabling the ecological transition. Data-driven approaches are increasingly employed in the management and planning of urban areas, contributing to a deeper understanding of health-related factors and supporting evidence-based decision-making.

Solutions based on these technologies are widely recognized, disseminated, and anticipated. However, their definition, deployment, and operation are not always executed optimally as each solution presents its unique challenges. This thesis addresses a specific issue that, under scientific rigor, has been analyzed to develop a feasible solution for the Smart Cities of today and the future.

The contributions detailed throughout the paper can be summarized as follows:

- **The IoT's role in measuring and correlating quantitative measurements and qualitative factors for architectural decision-making that affects human well-being in and around buildings (Chapter 2)**

  The essential role in measuring and monitoring environmental variables that influence human well-being is played by the IoT. However, it is recognized that environmental measurement is not a straightforward task, as a specific state of well-being is determined by various variables that are highly correlated.

  This chapter introduces environments of two types where the digitization of several variables is facilitated by IoT, permitting the analysis of these complex surroundings. An indoor monitoring system, PROMET&O, is presented, along with an outdoor monitoring system, in which the monitoring and validation of pollutant behaviors, such as particulate matter, and its interaction with architectural environments are facilitated by an IoT framework.

  It is revealed by this study that high spatiotemporal resolutions for continuous monitoring are required for concepts like well-being, in order to digitalize and quantify the health level of a given environment. This emphasizes the critical role played by IoT networks and advanced environmental simulations in promoting data-driven design, engaging users through data visualization, and ultimately enhancing the indoor and outdoor environmental quality.

- **An appropriate sampling rate that balances data accuracy and energy efficiency duty-cycle without sacrificing the quality of particulate matter concentration acquired by low-cost sensors. (Chapter 3)**

  Implementing an IoT solution is not a simple task, as each phenomenon presents its unique challenges, from the characteristics of the variables to be analyzed to the constraints inherent in IoT devices. In terms of data acquisition, this chapter introduces outdoor particulate matter monitoring as a case study. Given the architectural complexity of cities, there exist variations in pollution levels and various pollution sources, impacting the dynamics of particulate matter concentration.

  On the other hand, the sensors are energy-demanding, which complicates battery-powered deployment.

  This chapter presents a frequency analysis that determines the bandwidth under normal conditions to comply with the Nyquist-Shannon sampling theorem,

reducing redundant measurements that consume resources. From this study, a set of sampling frequencies is presented that can be applied depending on the type of dynamics one wishes to measure.

Under the obtained frequencies, the use of duty cycle to reduce sensor energy consumption and its effect on measurement precision is explored. It identifies points where pollution dynamics are lower, and the use of the duty cycle affects measurement precision to a lesser degree.

- **A novel hybrid model that integrates diverse deployment scenarios using a unified architectural framework and a standardized particulate matter monitoring station. This model harnesses a range of low-cost wireless technologies, thereby achieving high spatio-temporal resolutions. (Chapter 5)**

  The transformative potential of the IoT in establishing smart cities depends largely on the strategic deployment model used to balance operationally viable solutions with valuable, actionable data.

  Extensive research has evaluated different deployment methodologies, focusing on cost-effectiveness, and spatio-temporal resolutions. Yet, a comprehensive solution that maximizes these objectives simultaneously has not been found due to technological and scenario-specific limitations.

  This study analyzes three main deployment scenarios: fixed, participatory, and mobile sensor networks. Each offers unique advantages but none provide a complete solution independently. Consequently, a novel hybrid model that combines these three scenarios using a standard particulate matter monitoring station is proposed.

  The system architecture uses wireless technologies on unlicensed frequency bands, ensuring broad participation in data collection. This, along with easy data centralization and management, contributes to cost-effectiveness and high resolutions. Its scalability and flexibility, using data exchange protocols like MQTT and REST, make it a significant advantage for future IoT deployments.

  This study's findings suggest new ways to leverage IoT's potential in urban settings. The real-time data could aid urban planners, policymakers, and researchers in creating healthier, sustainable environments, thus offering promising opportunities for environmental research and urban development.

- **A framework to improve data quality, thereby increasing the coherence and reliability of IoT platforms that use low-cost light scattering sensors for air pollution monitoring (Chapter 5)**

  The value of an IoT solution is predominantly based on the quality of its data. Thus, the data received primarily from low-cost sensors must be processed and analyzed. These sensors often provide noisy, uncalibrated measurements, and in some instances, exhibit high failure rates.

  Anomaly detection, the identification of sensor malfunctions and failures that significantly impact the performance and reliability of a sensor network, is the focus of this chapter.

  A methodological pipeline is proposed, with each stage specifically designed to detect distinct types of anomalies. This comprehensive framework is characterized by the utilization of a combination of algorithms to detect prevalent sensor failures, exclude outliers, and calibrate sensors.

  More accurate measurements are yielded by the system, demonstrating resilience to anomalies resulting from partial or complete sensor failure. Additionally, a decrease in measurement dispersion is achieved, thereby enhancing the consistency in these sensor types compared to other reference calibration systems. The system also allows for the median of the measurements to remain within the manufacturer-specified error levels.

  In a broader context, the promising results obtained point towards a positive future for such sensors in measuring particulate matter within a smart city environment. The process of fault detection leads to proactive maintenance or replacement strategies, thereby ensuring prolonged sensor longevity. In stable climatic conditions, such as those near the Earth's equator, continuous monitoring is made possible by this approach. This feature differentiates it from cities encountering critical conditions at specific times of the year, as has been discussed in this research.

- **A Blockchain network architecture proposal for a national e-ID system with iris and fingerprint recognition features. The design, implementation, and validation of a Blockchain network for the proposed e-ID system through a new consensus method called tournament consensus algorithm (TCA). (Chapter 6)**

This chapter start with an examination of a private Blockchain Technology architecture, aimed at optimizing power usage and transactional efficiency in the context of general government services where security cannot be compromised. As such, the Tournament Consensus Algorithm (TCA) is employed in a private Blockchain deployment, satisfying the aforementioned requirements.

Further, the performance of the TCA is evaluated in a controlled setting, demonstrating its effectiveness across different transaction volumes. The results suggest that TCA requires an appropriate level of computing resources, yet it doesn't compromise the security integral to the chain's integrity. This stands in contrast to evaluations of Proof of Work and Proof of Stake algorithms, which exhibit greater power consumption and security concerns in private blockchains.

The proposed architecture suggests that the initiation of a blockchain transaction is primarily facilitated by smartcards. A robust authentication mechanism is proposed, integrating the three user authentication factors founded on the MoC mechanism prevalent in current smartcards and biometric systems. This approach effectively deters identity theft of document owners, while ensuring control and access to e-government services is maintained solely by the legitimate owner.

- **A hybrid architecture for data storage utilizes both centralized and decentralized storage resources, with data integrity safeguarded through the use of interlaced hashes. (Chapter 6)**

  The fusion of technologies like blockchain and IoT presents multiple advantages, particularly in protecting data integrity. This chapters ends outlining an architecture that exemplifies the synergistic potential of these technologies, introducing an enhanced layer of trust, transparency, and robust security to ensure data reliability and resistance to manipulation.

  Blockchain's integration is a pivotal advancement, bolstering user confidence in data authenticity. The architecture's resilience and reliability are further fortified by the immutable metadata stored in the blockchain and an added layer of data security provided by a dual-storage mechanism in the public blockchain. This mechanism deters data manipulation by linking two unique hashes, giving users the necessary data for calculating each hash.

Although Ethereum's Proof of Stake algorithm offers energy-efficient data integrity, continual exploration of alternatives to improve this efficiency is essential. Such progression must balance data integrity with sustainability and efficiency considerations. Therefore, this architecture could potentially inform the development of future secure, resilient, and reliable IoT systems, setting new benchmarks for data integrity.

- **A novel IoT system is proposed, which utilizes facial expression recognition and physiological data analysis to assist teachers in evaluating teaching effectiveness by determining students' level of attention during both in-person and remote lectures.(Chapter 7)**

  Building upon the heuristic designed in the prior case study, another complex issue is encountered, where the variable is not straightforwardly ascertainable. Therefore, an extrapolation grounded in the earlier analyses is undertaken. This chapter focuses on the use of dual, redundant systems for assessing student attention levels during in-person and remote lectures, particularly when no ground truth is available. Two central technical contributions are presented: the use of a Convolutional Neural Network (CNN) for facial expression recognition and a deterministic algorithm to examine physiological responses.

  These methodologies, both widely acknowledged, are combined in a unique, redundant manner to integrate data from facial recognition and physiological analyses. The CNN classifies emotions, while the physiological algorithm distinguishes between attention and distraction states. Extensive efforts were made to produce 'attention arrays' to facilitate the combined data's analysis and comparison.

  The proposed IoT system could aid teachers in gauging classroom attention levels and improving teaching quality by pinpointing key attention-required moments. Importantly, the system does not produce individual-specific attention data, and pre-merge data should not be shared to ensure student privacy.

  The study found a significant presence of 'Neutrality' outcomes, reducing the reliability of Neural Network results compared to Physiological Data analysis. Improved Neural Network performance and the addition of other data sources like blink counters and eye gaze direction controllers could enhance accuracy, provide additional insight into attention levels, and reinforce the system's redundancy.

References

[1] European Commission. Directorate General for the Information Society and Media. *Vision and challenges for realising the Internet of things.* Publications Office, 2010.

[2] Mckinsey Global Institute. The internet of thins: Mapping the value beyond the hype. https://www.mckinsey.com/~/media/mckinsey/industries/technology%20media%20and%20telecommunications/high%20tech/ou%20insights/the%20internet%20of%20things%20the%20value%20of%20digitizing%20the%20physical%20world/the-internet-of-things-mapping-the-value-beyond-the-hype.ashx, 2015.

[3] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer Networks*, 54(15):2787–2805, October 2010.

[4] International Telecommunication Union. The affordability of ICT services 2022. https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2022/ITU_Price_Brief_2022.pdf, 2022.

[5] Antti Martikkala, Joe David, Andrei Lobov, Minna Lanz, and Iñigo Flores Ituarte. Trends for low-cost and open-source IoT solutions development for industry 4.0. *Procedia Manufacturing*, 55:298–305, 2021.

[6] Shancang Li, Li Da Xu, and Shanshan Zhao. The internet of things: a survey. *Information Systems Frontiers*, 17(2):243–259, April 2014.

[7] Hans Schaffers, Nicos Komninos, Marc Pallot, Brigitte Trousse, Michael Nilsson, and Alvaro Oliveira. *Smart cities and the future internet: Towards cooperation frameworks for open innovation*. Springer Berlin Heidelberg, 2011.

[8] Andrea Zanella, Nicola Bui, Angelo Castellani, Lorenzo Vangelista, and Michele Zorzi. Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1):22–32, February 2014.

[9] Dana Cuff, Mark Hansen, and Jerry Kang. Urban sensing: out of the woods. *Communications of the ACM*, 51(3):24–33, 2008.

[10] Teemu Nuortio, Jari Kytöjoki, Harri Niska, and Olli Bräysy. Improved route planning and scheduling of waste collection and transport. *Expert systems with applications*, 30(2):223–232, 2006.

[11] Nicolas Maisonneuve, Matthias Stevens, Maria E Niessen, Peter Hanappe, and Luc Steels. Citizen noise pollution monitoring. *Association for Computing Machinery*, 2009.

[12] Wolfgang Kastner, Georg Neugschwandtner, Stefan Soucek, and H Michael Newman. Communication systems for building automation and control. *Proceedings of the IEEE*, 93(6):1178–1203, 2005.

[13] Bartolomeo Montrucchio, Edoardo Giusto, Mohammad Ghazi Vakili, Stefano Quer, Renato Ferrero, and Claudio Fornaro. A densely-deployed, high sampling rate, open-source air pollution monitoring wsn. *IEEE Transactions on Vehicular Technology*, 69(12):15786–15799, 2020.

[14] Ibrar Yaqoob, Ejaz Ahmed, Ibrahim Abaker Targio Hashem, Abdelmuttlib Ibrahim Abdalla Ahmed, Abdullah Gani, Muhammad Imran, and Mohsen Guizani. Internet of things architecture: Recent advances, taxonomy, requirements, and open challenges. *IEEE Wireless Communications*, 24(3):10–16, June 2017.

[15] Yasir Mehmood, Farhan Ahmad, Ibrar Yaqoob, Asma Adnane, Muhammad Imran, and Sghaier Guizani. Internet-of-things-based smart cities: Recent advances and challenges. *IEEE Communications Magazine*, 55(9):16–24, 2017.

[16] Francesco Concas, Julien Mineraud, Eemil Lagerspetz, Samu Varjonen, Xiaoli Liu, Kai Puolamäki, Petteri Nurmi, and Sasu Tarkoma. Low-cost outdoor air quality monitoring and sensor calibration. *ACM Transactions on Sensor Networks*, 17(2):1–44, May 2021.

[17] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE communications surveys & tutorials*, 17(4):2347–2376, 2015.

[18] Wafaa Anani, Abdelkader Ouda, and Ali Hamou. A survey of wireless communications for iot echo-systems. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, pages 1–6. IEEE, 2019.

[19] Lidia Morawska, Phong K. Thai, Xiaoting Liu, Akwasi Asumadu-Sakyi, Godwin Ayoko, Alena Bartonova, Andrea Bedini, Fahe Chai, Bryce Christensen, Matthew Dunbabin, Jian Gao, Gayle S.W. Hagler, Rohan Jayaratne, Prashant Kumar, Alexis K.H. Lau, Peter K.K. Louie, Mandana Mazaheri, Zhi Ning, Nunzio Motta, Ben Mullins, Md Mahmudur Rahman, Zoran Ristovski, Mahnaz Shafiei, Dian Tjondronegoro, Dane Westerdahl, and Ron Williams. Applications of low-cost sensing technologies for air quality monitoring and

exposure assessment: How far have they gone? *Environment International*, 116(April):286–299, 2018.

[20] Rajeev Kumar and M P S Bhatia. A systematic review of the security in cloud computing: Data integrity, confidentiality and availability. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 334–337, 2020.

[21] Bahar Farahani, Farshad Firouzi, and Markus Luecking. The convergence of iot and distributed ledger technologies (dlt): Opportunities, challenges, and solutions. *Journal of Network and Computer Applications*, 177:102936, 2021.

[22] Johannes Sedlmeir, Hans Ulrich Buhl, Gilbert Fridgen, and Robert Keller. The energy consumption of blockchain technology: Beyond myth. *Business & Information Systems Engineering*, 62(6):599–608, 2020.

[23] Department of Economic United Nations and Social Affairs Sustainable Development. Transforming our world: the 2030 agenda for sustainable development, 2015.

[24] Riccardo Giovanardi, Matteo; Trane, Matteo; Giusto, Edoardo; Ramirez Espinosa, Gustavo Adolfo; Chiavassa, Pietro; Montrucchio, Bartolomeo; Pollo. Promoting Healthy City through ICT technologies. An experience of Particulate Matter low-cost monitoring. In *36th Conference PLEA 2022. Sustainable Architecture and Urban Design*, pages 556–561. PLEA, 2022.

[25] Arianna Astolfi, Alessio Carullo, Virginia Fissore, Giuseppina Puglisi, Giuseppina Arcamone, Louena Shtrepi, Erica Raviola, Alberto Barbaro, Gustavo Ramirez Espinosa, Pietro Chiavassa, Edoardo Giusto, Gabriele Piccablotto, Fabio Saba, Davide Paesante, Giovanni Durando, Alice Lorenzati, Stefano Fantucci, Antonio Servetti, Bartolomeo Montrucchio, Franco Fiori, Anna Pellegrino, Vincenzo Corrado, Simona Paduos, Nicolas Sassoli, and Jana Clerici. Development and metrological characterization of a Multi-Sensor device for indoor environmental quality (IEQ) monitoring. In *2023 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv) (2023 IEEE MetroLivEnv)*, page 6, Milano, Italy, May 2023.

[26] Huading Lou and Dayi Ou. A comparative field study of indoor environmental quality in two types of open-plan offices: Open-plan administrative offices and open-plan research offices. *Building and Environment*, 148:394–404, 2019.

[27] Ioannis A Sakellaris, Dikaia E Saraga, Corinne Mandin, Célina Roda, Serena Fossati, Yvonne De Kluizenaar, Paolo Carrer, Sani Dimitroulopoulou, Victor G Mihucz, Tamás Szigeti, et al. Perceived indoor environment and occupants' comfort in european "modern" office buildings: The officair study. *International journal of environmental research and public health*, 13(5):444, 2016.

[28] Akira Tiele, Siavash Esfahani, James Covington, et al. Design and development of a low-cost, portable monitoring device for indoor environment quality. *Journal of Sensors*, 2018, 2018.

[29] Thomas Parkinson, Alex Parkinson, and Richard de Dear. Continuous ieq monitoring system: Context and development. *Building and Environment*, 149:15–25, 2019.

[30] Niels Lassen, Terje Josefsen, and Francesco Goia. Design and in-field testing of a multi-level system for continuous subjective occupant feedback on indoor climate. *Building and Environment*, 189:107535, 2021.

[31] Eziaku Onyeizu Rasheed and Hugh Byrd. Can self-evaluation measure the effect of ieq on productivity? a review of literature. *Facilities*, 35(11/12):601–621, 2017.

[32] Chien-fei Chen, Tianzhen Hong, Gerardo Zarazua de Rubens, Selin Yilmaz, Karol Bandurski, Zsófia Deme Bélafi, Marilena De Simone, Mateus Vinícius Bavaresco, Yu Wang, Pei-ling Liu, et al. Culture, conformity, and carbon? a multi-country analysis of heating and cooling practices in office buildings. *Energy Research & Social Science*, 61:101344, 2020.

[33] Wenjuan Wei, Pawel Wargocki, Johann Zirngibl, Jana Bendžalová, and Corinne Mandin. Review of parameters used to assess the quality of the indoor environment in green building certification schemes for offices and hotels. *Energy and buildings*, 209:109683, 2020.

[34] E220-19. test method for calibration of thermocouples by comparison techniques. https://doi.org/10.1520/e0220-19.

[35] Pietro Chiavassa, Filippo Gandino, and Edoardo Giusto. An investigation on duty-cycle for particulate matter monitoring with light-scattering sensors. In *2021 6th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6. IEEE, 2021.

[36] Gustavo Ramirez Espinosa, Bartolomeo Montrucchio, Edoardo Giusto, and Maurizio Rebaudengo. Low-cost pm sensor behaviour based on duty-cycle analysis. In *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–8. IEEE, 2021.

[37] International Organization for Standardization. Determination of particle size distribution, single particle light interaction methods: Détermination de la distribution granulomètrique, méthodes d'interaction lumineuse de paritules uniques.

[38] Erika Brattich, Alessandro Bracci, Alessandro Zappi, Pietro Morozzi, Silvana Di Sabatino, Federico Porcù, Francesca Di Nicola, and Laura Tositti. How to get the best from low-cost particulate matter sensors: Guidelines and practical recommendations. *Sensors*, 20(11):3073, 2020.

[39] Di Liu, Qiang Zhang, Jingkun Jiang, and Da Ren Chen. Performance calibration of low-cost and portable particular matter (PM) sensors. *Journal of Aerosol Science*, 112(March):1–10, 2017.

[40] World Health Organization. *World health statistics 2021: monitoring health for the SDGs, sustainable development goals*. World Health Organization, 2021.

[41] Air quality in europe - 2020 report. https://www.eea.europa.eu/publications/air-quality-in-europe-2020-report, Oct 2020.

[42] Xiao Wu, Rachel C Nethery, M Benjamin Sabath, Danielle Braun, and Francesca Dominici. Exposure to air pollution and COVID-19 mortality in the united states: A nationwide cross-sectional study. *medRxiv*, April 2020.

[43] Riccardo Pollo, Elisa Biolchini, Giulia Squillacioti, and Roberto Bono. Designing the healthy city: An interdisciplinary approach. In *Sustainable Mediterranean Construction, 11*, pages 150–155, 12 2020.

[44] Luigi Atzori, Claudia Campolo, Bin Da, Roberto Girau, Antonio Iera, Giacomo Morabito, and Salvatore Quattropani. Smart devices in the social loops: Criteria and algorithms for the creation of the social links. *Future Generation Computer Systems*, 97:327–339, 2019.

[45] Simone Mora, Amin Anjomshoaa, Tom Benson, Fábio Duarte, and Carlo Ratti. Towards large-scale drive-by sensing with multi-purpose city scanner nodes. In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 743–748, 2019.

[46] Frances Kane, Joseph Abbate, Eric C. Landahl, and Mark J. Potosnak. Monitoring particulate matter with wearable sensors and the influence on student environmental attitudes. *Sensors*, 22(3), 2022.

[47] Joost Wesseling, Wouter Hendricx, Henri de Ruiter, Sjoerd van Ratingen, Derko Drukker, Maaike Huitema, Claar Schouwenaar, Geert Janssen, Stephen van Aken, Jan Smeenk, Arjen Hof, and Erik Tielemans. Assessment of PM2.5 exposure during cycle trips in the netherlands using low-cost sensors. *International Journal of Environmental Research and Public Health*, 18(11):6007, June 2021.

[48] M. Giovanardi, E. Giusto, and R. Pollo. Digital infrastructures in facade components for building management. In *Design in the digital age. Technology, Nature, Culture, Sant'Arcangelo di Romagna:Maggioli.*, pages 243–248, 2021.

[49] Massimiliano Viglioglia, Matteo Giovanardi, Riccardo Pollo, and Pier Paolo Peruccio. Smart district and circular economy: The role of ict solutions in promoting circular cities. *Sustainability*, 13(21), 2021.

[50] Riccardo Pollo, Matteo Trane, and Matteo Giovanardi. Urban metabolism, interdisciplinary models and design at micro-urban scale. *TECHNE - Journal of Technology for Architecture and Environment, 21*, page 154–164, Mar. 2021.

[51] Ronald Kellett, Andreas Christen, Nicholas C. Coops, Michael van der Laan, Ben Crawford, Thoreau Rory Tooke, and Inna Olchovski. A systems approach to carbon cycling and emissions modeling at an urban neighborhood scale. *Landscape and Urban Planning*, 110:48–58, 2013.

[52] Liyan Rui, Riccardo Buccolieri, Zhi Gao, Elisa Gatto, and Wowo Ding. Study of the effect of green quantity and structure on thermal comfort and air quality in an urban-like residential district by ENVI-met modelling. *Building Simulation*, 12(2):183–194, December 2018.

[53] Kerstin Krellenberg and Florian Koch. Conceptualizing interactions between SDGs and urban sustainability transformations in covid-19 times. *Politics and Governance*, 9(1):200–210, February 2021.

[54] Jordi Honey-Rosés, Isabelle Anguelovski, Vincent K. Chireh, Carolyn Daher, Cecil Konijnendijk van den Bosch, Jill S. Litt, Vrushti Mawani, Michael K. McCall, Arturo Orellana, Emilia Oscilowicz, Ulises Sánchez, Maged Senbel, Xueqi Tan, Erick Villagomez, Oscar Zapata, and Mark J Nieuwenhuijsen. The impact of COVID-19 on public space: an early review of the emerging questions design, perceptions and inequities. *Cities Health*, 5(sup1):S263–S279, July 2020.

[55] Oriana Ramirez-Rubio, Carolyn Daher, Gonzalo Fanjul, Mireia Gascon, Natalie Mueller, Leire Pajín, Antoni Plasencia, David Rojas-Rueda, Meelan Thondoo, and Mark J. Nieuwenhuijsen. Urban health: an example of a "health in all policies" approach in the context of SDGs implementation. *Globalization and Health*, 15(1), December 2019.

[56] Gustavo Ramirez Espinosa, Bartolomeo Montrucchio, Filippo Gandino, and Maurizio Rebaudengo. Frequency Analysis of Particulate Matter in Urban Environments under Low-cost Sensors. In *2021 International Conference on Computer Communication and Artificial Intelligence (CCAI)*, pages 97–105. IEEE, 5 2021.

[57] World Bank. Urban Population statistics - Urban population (% of Total). https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS, 2016.

[58] World Health Organization. Ambient (outdoor) air pollution. https://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health, 2018.

[59] A. Dobre, S. J. Arnold, R. J. Smalley, J. W.D. Boddy, J. F. Barlow, A. S. Tomlin, and S. E. Belcher. MegaSense: Feasibility of Low-Cost Sensors for Pollution Hot-spot Detection. *IEEE International Conference on Industrial Informatics (INDIN)*, 2019-July:1083–1090, 2019.

[60] A. Dobre, S. J. Arnold, R. J. Smalley, J. W.D. Boddy, J. F. Barlow, A. S. Tomlin, and S. E. Belcher. Flow field measurements in the proximity of an urban intersection in London, UK. *Atmospheric Environment*, 39(26):4647–4657, 2005.

[61] Matthias Budde, Rayan El Masri, Till Riedel, and Michael Beigl. Enabling low-cost particulate matter measurement for participatory sensing scenarios. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia, MUM 2013*, 2013.

[62] Aakash C. Rai, Prashant Kumar, Francesco Pilla, Andreas N. Skouloudis, Silvana Di Sabatino, Carlo Ratti, Ansar Yasar, and David Rickerby. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Science of the Total Environment*, 607-608:691–705, 2017.

[63] C. Borrego, A. M. Costa, J. Ginja, M. Amorim, M. Coutinho, K. Karatzas, Th Sioumis, N. Katsifarakis, K. Konstantinidis, S. De Vito, E. Esposito, P. Smith, N. André, P. Gérard, L. A. Francis, N. Castell, P. Schneider, M. Viana, M. C. Minguillón, W. Reimringer, R. P. Otjes, O. von Sicard, R. Pohle, B. Elen, D. Suriano, V. Pfister, M. Prato, S. Dipinto, and M. Penza. Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise. *Atmospheric Environment*, 147(2):246–263, 2016.

[64] Nuria Castell, Franck R. Dauge, Philipp Schneider, Matthias Vogt, Uri Lerner, Barak Fishbain, David Broday, and Alena Bartonova. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International*, 99:293–302, 2017.

[65] Balz Maag, Zimu Zhou, and Lothar Thiele. A Survey on Sensor Calibration in Air Pollution Monitoring Deployments. *IEEE Internet of Things Journal*, 5(6):4857–4870, 2018.

[66] C Borrego, M Coutinho, A M Costa, J Ginja, C Ribeiro, A Monteiro, I Ribeiro, J Valente, J H Amorim, H Martins, D Lopes, and A I Miranda. Challenges for a New Air Quality Directive: The role of monitoring and modelling techniques. *Urban Climate*, 14:328–341, 2015.

[67] European Union. CITI-SENSE : Development of citizen observatories for improvement of quality of life in cities. http://www.citi-sense.eu/.

[68] Jonathan E. Thompson. Crowd-sourced air quality studies: A review of the literature & portable sensors. *Trends in Environmental Analytical Chemistry*, 11:23–34, jul 2016.

[69] Wei Yi, Kin Lo, Terrence Mak, Kwong Leung, Yee Leung, and Mei Meng. A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems. *Sensors*, 15(12):31392–31427, dec 2015.

[70] Vana Jelicic, Michele Magno, Davide Brunelli, Giacomo Paci, and Luca Benini. Context-adaptive multimodal wireless sensor network for energy-efficient gas monitoring. *IEEE Sensors Journal*, 13(1):328–338, 2013.

[71] B. Montrucchio, E. Giusto, M. Ghazi, S. Quer, R. Ferrero, and C. Fornaro. A Densely-Deployed, High Sampling Rate, Open-Source Air Pollution Monitoring WSN | IEEE DataPort. https://ieee-dataport.org/open-access/densely-deployed-high-sampling-rate-open-source-air-pollution-monitoring-wsn.

[72] W Jiao, G Hagler, R Williams, R Sharpe, R Brown, D Garver, R Judge, M Caudill, J Rickard, M Davis, L Weinstock, S Zimmer-Dauphinee, and K Buckley. Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmospheric Measurement Techniques*, 9(11):5281–5292, 2016.

[73] K K Johnson, M H Bergin, A G Russell, and G S W Hagler. Using Low Cost Sensors to Measure Ambient Particulate Matter Concentrations and On-Road Emissions Factors. *Atmospheric Measurement Techniques Discussions*, 2016:1–22, 2016.

[74] Meiling Gao, Junji Cao, and Edmund Seto. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. *Environmental Pollution*, 199:56–65, 4 2015.

[75] Nicholas A.S. Hamm, Marco van Lochem, Gerard Hoek, René Otjes, Sandra van der Sterren, and Hans Verhoeven. "The Invisible Made Visible": Science and Technology. In *AiREAS: Sustainocracy for a Healthy City. SpringerBriefs on Case Studies of Sustainable Development*, pages 51–77. Springer, Cham, 2016.

[76] Anondo Mukherjee, Levi Stanton, Ashley Graham, and Paul Roberts. Assessing the Utility of Low-Cost Particulate Matter Sensors over a 12-Week Period in the Cuyama Valley of California. *Sensors*, 17(8):1805, 8 2017.

[77] G. Olivares and S. Edwards. The Outdoor Dust Information Node (ODIN) – development and performance assessment of a low cost ambient dust sensor. *Atmospheric Measurement Techniques Discussions*, 8(7):7511–7533, 7 2015.

[78] H. Ali, J. K. Soe, and Steven R. Weller. A real-time ambient air quality monitoring wireless sensor network for schools in smart cities. In *2015 IEEE 1st International Smart Cities Conference, ISC2 2015*. Institute of Electrical and Electronics Engineers Inc., 12 2015.

[79] Nadezda Zikova, Philip K. Hopke, and Andrea R. Ferro. Evaluation of new low-cost particle monitors for pm2.5 concentrations measurements. *Journal of Aerosol Science*, 105:24 – 34, 2017.

[80] K. E. Kelly, J. Whitaker, A. Petty, C. Widmer, A. Dybwad, D. Sleeth, R. Martin, and A. Butterfield. Ambient and laboratory evaluation of a low-cost particulate matter sensor. *Environmental Pollution*, 221:491–500, 2 2017.

[81] Leigh R. Crilley, Marvin Shaw, Ryan Pound, Louisa J. Kramer, Robin Price, Stuart Young, Alastair C. Lewis, and Francis D. Pope. Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmospheric Measurement Techniques*, 11(2):709–720, 2 2018.

[82] E. Giusto, R. Ferrero, F. Gandino, B. Montrucchio, M. Rebaudengo, and M. Zhang. Particulate matter monitoring in mixed indoor/outdoor industrial applications: A case study. In *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, volume 1, pages 838–844, 2018.

[83] D. K. Arvind, Janek Mann, Andrew Bates, and Konstantin Kotsev. The AirSpeck Family of Static and Mobile Wireless Air Quality Monitors. In *Proceedings - 19th Euromicro Conference on Digital System Design, DSD 2016*, pages 207–214. Institute of Electrical and Electronics Engineers Inc., 10 2016.

[84] Yan Zhuang, Feng Lin, Yoo Eun-Hye, and Wenyao Xu. AirSense: A portable context-sensing device for personal air quality monitoring. *MobileHealth 2015 - Proceedings of the 2015 Workshop on Pervasive Wireless Healthcare, co-located with MobiHoc 2015*, pages 17–22, 2015.

[85] Ruizhe Zhang, Daniele Ravi, Guang Zhong Yang, and Benny Lo. A personalized air quality sensing system-A preliminary study on assessing the air quality of London underground stations. In *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2017*, pages 111–114. Institute of Electrical and Electronics Engineers Inc., 5 2017.

[86] Miguel Alvarado, Felipe Gonzalez, Andrew Fletcher, and Ashray Doshi. Towards the Development of a Low Cost Airborne Sensing System to Monitor Dust Particles after Blasting at Open-Pit Mine Sites. *Sensors*, 15(8):19667–19687, 8 2015.

[87] Man Sing Wong, Tsan Pong Yip, and Esmond Mok. Development of a personal integrated environmental monitoring system. *Sensors (Switzerland)*, 14(11):22065–22081, 11 2014.

[88] X. Liu, B. Li, A. Jiang, S. Qi, C. Xiang, and N. Xu. A bicycle-borne sensor for monitoring air pollution near roadways. In *2015 IEEE International Conference on Consumer Electronics - Taiwan, ICCE-TW 2015*, pages 166–167, 2015. Cited By :15.

[89] D M Holstius, A Pillarisetti, K R Smith, and E Seto. Field calibrations of a low-cost aerosol sensor Atmospheric Measurement Techniques Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in

California Field calibrations of a low-cost aerosol sensor. *Atmos. Meas. Tech. Discuss*, 7:605–632, 2014.

[90] Yun Cheng, Xiucheng Li, Zhijun Li, Shouxu Jiang, Yilong Li, Ji Jia, and Xiaofan Jiang. AirCloud: a cloud-based air-quality monitoring system for everyone. *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems - SenSys '14*, pages 251–265, 2014.

[91] Rita T. Tse and Yubin Xiao. A portable Wireless Sensor Network system for real-time environmental monitoring. In *WoWMoM 2016 - 17th International Symposium on a World of Wireless, Mobile and Multimedia Networks*. Institute of Electrical and Electronics Engineers Inc., 7 2016.

[92] Tingting Cao and Jonathan E. Thompson. Portable, Ambient PM2.5 Sensor for Human and/or Animal Exposure Studies. *Analytical Letters*, 50(4):712–723, 3 2017.

[93] Susanne Steinle, Stefan Reis, Clive E. Sabel, Sean Semple, Marsailidh M. Twigg, Christine F. Braban, Sarah R. Leeson, Mathew R. Heal, David Harrison, Chun Lin, and Hao Wu. Personal exposure monitoring of PM2.5 in indoor and outdoor microenvironments. *Science of the Total Environment*, 508:383–394, 3 2015.

[94] Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. Real-time air quality monitoring through mobile sensing in metropolitan areas. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 1, New York, New York, USA, 2013. ACM Press.

[95] Yi Gao, Wei Dong, Kai Guo, Xue Liu, Yuan Chen, Xiaojin Liu, Jiajun Bu, and Chun Chen. Mosaic: A low-cost mobile sensing system for urban air quality monitoring. *Proceedings - IEEE INFOCOM*, 2016-July, 2016.

[96] David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin, Martin Fierz, and Lothar Thiele. Pushing the spatio-temporal resolution limit of urban air pollution maps. In *2014 IEEE International Conference on Pervasive Computing and Communications, PerCom 2014*, pages 69–77. IEEE, 3 2014.

[97] Samer Mansour, Nidal Nasser, Lutful Karim, and Asmaa Ali. Wireless sensor network-based air quality monitoring system. *2014 International Conference on Computing, Networking and Communications, ICNC 2014*, pages 545–550, 2014.

[98] Rundong Tian, Christine Dierk, Christopher Myers, and Eric Paulos. MyPart: Personal, Portable, Accurate, Airborne Particle Counting. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1338–1348, New York, NY, USA, may 2016. ACM.

[99] Mayukh Roy Chowdhury, Swades De, Narendra Kumar Shukla, and Ranendra N. Biswas. Energy-Efficient Air Pollution Monitoring with Optimum Duty-Cycling on a Sensor Hub. In *2018 Twenty Fourth National Conference on Communications (NCC)*, pages 1–6. IEEE, feb 2018.

[100] O. Tchepel and C. Borrego. Frequency analysis of air quality time series for traffic related pollutants. *Journal of Environmental Monitoring*, 12(2):544–550, 2010.

[101] The World Air Quality Project. Rubino, Torino, Piemonte, Italy Air Pollution: Real-time Air Quality Index. https://aqicn.org/city/italy/piemonte/torino/rubino/.

[102] Honeywell International Inc. HPMA115S0 Particulate Matter Sensors - Honeywell. https://sensing.honeywell.com/hpma115s0-xxx-particulate-matter-sensors.

[103] Exact Signal Measurements using FFT Analysis. http://nbn-resolving.de/urn:nbn:de:hbz:386-kluedo-42930, 2016.

[104] U Henriksson. Power spectrum and bandwidth. https://www.commsys.isy.liu.se/TSDT45/Material/UlfsSpectrum2003.pdf, 2003.

[105] Jean Bousquet, Josep M. Anto, Isabella Annesi-Maesano, Toni Dedeu, Eve Dupas, Jean Louis Pépin, Landry Stephane Zeng Eyindanga, Sylvie Arnavielhe, Julia Ayache, Xavier Basagana, Samuel Benveniste, Nuria Calves Venturos, Hing Kin Chan, Mehdi Cheraitia, Yves Dauvilliers, Judith Garcia-Aymerich, Ingrid Jullian-Desayes, Chitra Dinesh, Daniel Laune, Jade Lu Dac, Ismael Nujurally, Giovanni Pau, Robert Picard, Xavier Rodo, Renaud Tamisier, Michael Bewick, Nils E. Billo, Wienczyslawa Czarlewski, Joao Fonseca, Ludger Klimek, Oliver Pfaar, and Jean Marc Bourez. POLLAR: Impact of air POLLution on Asthma and Rhinitis; A European Institute of Innovation and Technology Health (EIT Health) project. *Clinical and Translational Allergy*, 8(1):36, 9 2018.

[106] EUR-Lex - 02008L0050-20150918 - EN - EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1486474738782&uri=CELEX:02008L0050-20150918.

[107] Report: A Guide for Local Authorities Purchasing Air Quality Monitoring Equipment - Defra, UK. https://uk-air.defra.gov.uk/library/reports?report_id=386.

[108] Città metropolitana di Torino. Turin municipality, environmental monitoring stations map. http://www.cittametropolitana.torino.it/cms/ambiente/qualita-aria/rete-monitoraggio/stazioni-monitoraggio.

[109] Milena Jovašević-Stojanović, Alena Bartonova, Dušan Topalović, Ivan Lazović, Boris Pokrić, and Zoran Ristovski. On the use of small and cheaper

sensors and devices for indicative citizen-based monitoring of respirable particulate matter. *Environmental Pollution*, 206:696–704, 2015.

[110] Giuseppe Anastasi, Marco Conti, Mario Di Francesco, and Andrea Passarella. Energy conservation in wireless sensor networks: A survey. *Ad Hoc Networks*, 7(3):537–568, 5 2009.

[111] Felicia Engmann, Ferdinand Apietu Katsriku, Jamal Deen Abdulai, Kofi Sarpong Adu-Manu, and Frank Kataka Banaseka. Prolonging the Lifetime of Wireless Sensor Networks: A Review of Current Techniques. *Wireless Communications and Mobile Computing*, 2018, 2018.

[112] Elena Austin, Igor Novosselov, Edmund Seto, and Michael G Yost. Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor. *PLoS ONE*, 10(9), 2015.

[113] Davide Aguiari, Giovanni Delnevo, Lorenzo Monti, Vittorio Ghini, Silvia Mirri, Paola Salomoni, Giovanni Pau, Marcus Im, Rita Tse, Mongkol Ekpanyapong, and Roberto Battistini. Canarin II: Designing a smart e-bike eco-system. *CCNC 2018 - 2018 15th IEEE Annual Consumer Communications and Networking Conference*, 2018-Janua(762013):1–6, 2018.

[114] Boris Dessimond, Isabella Annesi-Maesano, Jean Louis Pepin, Salim Srairi, and Giovanni Pau. Academically produced air pollution sensors for personal exposure assessment: The canarin project. *Sensors*, 21(5):1–18, 3 2021.

[115] Vana Jelicic, Michele Magno, Davide Brunelli, Giacomo Paci, and Luca Benini. Context-adaptive multimodal wireless sensor network for energy-efficient gas monitoring. *IEEE Sensors Journal*, 13(1):328–338, 2013.

[116] Kavi K. Khedo, Rajiv Perseedoss, Avinash Mungur, University of Mauritius, and Mauritius. A wireless sensor network air pollution monitoring system. *International Journal of Wireless & Mobile Networks*, 2(2):31–45, 5 2010.

[117] Juan Botero-Valencia, Luis Castano-Londono, David Marquez-Viloria, and Mateo Rico-Garcia. Data reduction in a low-cost environmental monitoring system based on LoRa for WSN. *IEEE Internet of Things Journal*, 6(2):3024–3030, 4 2019.

[118] FiPy - Pycom -Five Network Development Board for IoT. https://docs.pycom.io/datasheets/development/fipy/.

[119] The Internet of Things with ESP32. http://esp32.net/.

[120] MicroPython - Python for microcontrollers. https://micropython.org/.

[121] Jason Jingshi Li, Boi Faltings, Olga Saukh, David Hasenfratz, and Jan Beutel. Sensing the air we breathe—the opensense zurich dataset. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, pages 323–325, 2012.

[122] David E Williams, Geoff S Henshaw, Mark Bart, Greer Laing, John Wagner, Simon Naisbitt, and Jennifer A Salmond. Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network. *Measurement Science and Technology*, 24(6):065803, 2013.

[123] Mark Bart, David E Williams, Bruce Ainslie, Ian McKendry, Jennifer Salmond, Stuart K Grange, Maryam Alavi-Shoshtari, Douw Steyn, and Geoff S Henshaw. High density ozone monitoring using gas sensitive semiconductor sensors in the lower fraser valley, british columbia. *Environmental science & technology*, 48(7):3970–3977, 2014.

[124] I Heimann, VB Bright, MW McLeod, MI Mead, OAM Popoola, GB Stewart, and RL Jones. Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. *Atmospheric Environment*, 113:10–19, 2015.

[125] Jahangir Ikram, Amer Tahir, Hasanat Kazmi, Zonia Khan, Rabi Javed, and Usama Masood. View: implementing low cost air quality monitoring solution for urban areas. *Environmental Systems Research*, 1:1–8, 2012.

[126] Sharon Moltchanov, Ilan Levy, Yael Etzion, Uri Lerner, David M Broday, and Barak Fishbain. On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Science of The Total Environment*, 502:537–547, 2015.

[127] Michael D Mueller, David Hasenfratz, Olga Saukh, Martin Fierz, and Christoph Hueglin. Statistical modelling of particle number concentration in zurich at high spatio-temporal resolution utilizing data from a mobile sensor network. *Atmospheric Environment*, 126:171–181, 2016.

[128] Domenico Suriano, Mario Prato, Valerio Pfister, Gennaro Cassano, Giuseppe Camporeale, Sebastiano Dipinto, Michele Penza, et al. 15-stationary and mobile low-cost gas sensor-systems for air quality monitoring applications. In *Fourth Scientific Meeting EuNetAir*, pages 55–58, 2015.

[129] Bart Elen, Jan Peters, Martine Van Poppel, Nico Bleux, Jan Theunis, Matteo Reggente, and Arnout Standaert. The aeroflex: a bicycle for mobile air quality measurements. *Sensors*, 13(1):221–240, 2012.

[130] Michael Jerrett, David Donaire-Gonzalez, Olalekan Popoola, Roderic Jones, Ronald C Cohen, Estela Almanza, Audrey De Nazelle, Iq Mead, Glòria Carrasco-Turigas, Tom Cole-Hunter, et al. Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science. *Environmental research*, 158:286–294, 2017.

[131] Kinza Shafique, Bilal A Khawaja, Farah Sabir, Sameer Qazi, and Muhammad Mustaqim. Internet of things (iot) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5g-iot scenarios. *Ieee Access*, 8:23022–23040, 2020.

[132] Rashmi Sharan Sinha, Yiqiao Wei, and Seung-Hoon Hwang. A survey on lpwa technology: Lora and nb-iot. *Ict Express*, 3(1):14–21, 2017.

[133] Ramon Sanchez-Iborra and Maria-Dolores Cano. State of the art in lp-wan solutions for industrial iot services. *Sensors*, 16(5):708, 2016.

[134] Kais Mekki, Eddy Bajic, Frederic Chaxel, and Fernand Meyer. A comparative study of lpwan technologies for large-scale iot deployment. *ICT Express*, 5(1):1–7, 2019.

[135] Jie Ding, Mahyar Nemati, Chathurika Ranaweera, and Jinho Choi. Iot connectivity technologies and applications: A survey. *IEEE Access*, PP:1–1, 04 2020.

[136] Laura Joris, François Dupont, Philippe Laurent, Pierre Bellier, Serguei Stoukatch, and Jean-Michel Redouté. An autonomous sigfox wireless sensor node for environmental monitoring. *IEEE Sensors Letters*, 3(7):01–04, 2019.

[137] Floarea Pitu and Nicoleta Cristina Gaitan. Surveillance of sigfox technology integrated with environmental monitoring. In *2020 International Conference on Development and Application Systems (DAS)*, pages 69–72, 2020.

[138] Niemah I. Osman and Esra B. Abbas. Simulation and modelling of lora and sigfox low power wide area network technologies. In *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, pages 1–5, 2018.

[139] Semtech Corporation. LoRa and LoRaWAN: A Technical Overview. https://lora-developers.semtech.com/documentation/tech-papers-and-guides/lora-and-lorawan/.

[140] Rachel Kufakunesu, Gerhard P. Hancke, and Adnan M. Abu-Mahfouz. A survey on adaptive data rate optimization in lorawan: Recent solutions and major challenges. *Sensors*, 20(18), 2020.

[141] Gonçalo Vicente and Gonçalo Marques. Air quality monitoring through lora technologies: A literature review. In *2020 International Conference on Decision Aid Sciences and Application (DASA)*, pages 350–354, 2020.

[142] Yohan Han and Jongpil Jeong. Air pollution measurement platform based on lora and blockchain for industrial iot applications. In *Computational Science and Its Applications–ICCSA 2020: 20th International Conference, Cagliari, Italy, July 1–4, 2020, Proceedings, Part II 20*, pages 613–622. Springer, 2020.

[143] Varada Raju, Addala Satya Narayana Varma, and Y Satyanarayana Raju. An environmental pollution monitoring system using lora. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pages 3521–3526, 2017.

[144] E Raghuveera, P Kanakaraja, K Hari Kishore, C Tanvi Sriya, Durga Prasad B, and B Sai Krishna Teja Lalith. An iot enabled air quality monitoring system using lora and lpwan. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 453–459, 2021.

[145] Mia Rosmiati, Moch Fachru Rizal, Fitri Susanti, and Gilang Fahreza Alfisyahrin. Air pollution monitoring system using lora modul as transceiver system. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 17(2):586–592, 2019.

[146] Thanpitcha Atiwanwong and Saweth Hongprasit. A low-power real-time pollution monitoring system using esp lora. *Engineering Access*, 6(1):36–40, 2020.

[147] Rizky Firdaus, Muhammad Ary Murti, and Ibnu Alinursafa. Air quality monitoring system based internet of things (iot) using lpwan lora. In *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, pages 195–200, 2019.

[148] Sujuan Liu, Chuyu Xia, and Zhenzhen Zhao. A low-power real-time air quality monitoring system using lpwan based on lora. In *2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, pages 379–381, 2016.

[149] Evariste Twahirwa, Kambombo Mtonga, Desire Ngabo, and Santhi Kumaran. A lora enabled iot-based air quality monitoring system for smart city. In *2021 IEEE World AI IoT Congress (AIIoT)*, pages 0379–0385, 2021.

[150] Supongmen Walling, Jayasree Sengupta, and Sipra Das Bit. A low-cost real-time iot based air pollution monitoring using lora. In *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pages 1–6, 2019.

[151] Sandra Sendra, Jose Luis Garcia-Navas, Pablo Romero-Diaz, and Jaime Lloret. Collaborative lora-based sensor network for pollution monitoring in smart cities. In *2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC)*, pages 318–323, 2019.

[152] Chao-Linag Hsieh, Zheng-Wei Ye, Chen-Kang Huang, Yeun-Chung Lee, Chih-Hong Sun, Tzai-Hung Wen, Jehn-Yih Juang, and Joe-Air Jiang. A vehicle monitoring system based on the lora technique. *International Journal of Transport and Vehicle Engineering*, 11(5):1100–1106, 2017.

[153] Liang-Yu Chen, Hsiao-Shih Huang, Cheng-Ju Wu, Yi-Ting Tsai, and Yue-Shan Chang. A lora-based air quality monitor on unmanned aerial vehicle for smart city. In *2018 International Conference on System Science and Engineering (ICSSE)*, pages 1–5, 2018.

[154] Rosa M Camarillo-Escobedo, Jorge L Flores-Nuñez, Juana M Camarillo-Escobedo, Pedro Marin-Montoya, J Ramón Amparan-Estrada, and G Garcia-Torales. Remote sensing system to monitoring of quality air using unmanned aerial vehicles and lora communication. In *Infrared Remote Sensing and Instrumentation XXIX*, volume 11830, pages 137–146. SPIE, 2021.

[155] Markus Knoll, Philipp Breitegger, and Alexander Bergmann. Low-power wide-area technologies as building block for smart sensors in air quality measurements. *Elektrotechnik und Informationstechnik*, 135(6):416–422, 2018.

[156] Wi-Fi Alliance. Wi-Fi Genarations. https://www.wi-fi.org/discover-wi-fi.

[157] The Working Group for WLAN Standards. IEEE 802.11 Wireless Local Area Networks. https://www.ieee802.org/11/.

[158] Andrii Polianytsia, Olena Starkova, and Kostiantyn Herasymenko. Survey of hardware iot platforms. In *2016 Third International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T)*, pages 152–153, 2016.

[159] Somayya Madakam, R. Ramaswamy, and Siddharth Tripathi. Internet of things (IoT): A literature review. *Journal of Computer and Communications*, 03(05):164–173, 2015.

[160] Farzad Samie, Lars Bauer, and Jörg Henkel. IoT technologies for embedded computing. In *Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis*. ACM, October 2016.

[161] Wi-Fi Alliance. Wi-Fi Certified 6. https://www.wi-fi.org/discover-wi-fi/wi-fi-certified-6.

[162] Serbulent Tozlu. Feasibility of wi-fi enabled sensors for internet of things. In *2011 7th International Wireless Communications and Mobile Computing Conference*, pages 291–296, 2011.

[163] Serbulent Tozlu, Murat Senel, Wei Mao, and Abtin Keshavarzian. Wi-fi enabled sensors for internet of things: A practical approach. *IEEE Communications Magazine*, 50(6):134–143, 2012.

[164] Hanseul Hong, Young Kim, and Ronny Kim. A low-power WLAN communication scheme for IoT WLAN devices using wake-up receivers. *Applied Sciences*, 8(1):72, January 2018.

[165] R. Piedrahita, Y. Xiang, N. Masson, J. Ortega, A. Collier, Y. Jiang, K. Li, R. P. Dick, Q. Lv, M. Hannigan, and L. Shang. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques*, 7(10):3325–3336, October 2014.

[166] IEEE Standard Association. IEEE 802.11ah-2016. https://standards.ieee.org/ieee/802.11ah/4960/.

[167] WiMax Forum. WiMax Forum. https://wimaxforum.org.

[168] IEEE Standard Association. IEEE 802.15.4-2020 - IEEE Standard for Low-Rate Wireless Networks. https://standards.ieee.org/ieee/802.15.4/7029/.

[169] ZigBee Alliance. ZigBee Specification. https://zigbeealliance.org/wp-content/uploads/2019/11/docs-05-3474-21-0csg-zigbee-specification.pdf.

[170] Internet Engineering Task Force. IPv6 over Low power WPAN. https://datatracker.ietf.org/wg/6lowpan/about/.

[171] Paolo Baronti, Prashant Pillai, Vince WC Chook, Stefano Chessa, Alberto Gotta, and Y Fun Hu. Wireless sensor networks: A survey on the state of the art and the 802.15. 4 and zigbee standards. *Computer communications*, 30(7):1655–1695, 2007.

[172] Yang Xiao, Hsiao-Hwa Chen, Bo Sun, Ruhai Wang, and Sakshi Sethi. Mac security and security overhead analysis in the ieee 802.15. 4 wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking*, 2006:1–12, 2006.

[173] Muhammad Sajjad Akbar, Zawar Hussain, Michael Sheng, and Rajan Shankaran. Wireless body area sensor networks: Survey of mac and routing protocols for patient monitoring under ieee 802.15. 4 and ieee 802.15. 6. *Sensors*, 22(21):8279, 2022.

[174] Stig Petersen, Bård Myhre, Paula Doyle, Erik Mikkelsen, Simon Carlsen, Dag Sjong, Amund Skavhaug, Jan Hendrik Van Der Linden, and Mark Sansom. A survey of wireless technology for the oil and gas industry. *SPE Projects, Facilities & Construction*, 3(04):1–8, 2008.

[175] Prativa P Saraswala. A survey on routing protocols in zigbee network. *International Journal of Engineering Science and Innovative Technology (IJESIT)*, 2(1):471–476, 2013.

[176] Mounib Khanafer, Mouhcine Guennoun, and Hussein T Mouftah. A survey of beacon-enabled ieee 802.15. 4 mac protocols in wireless sensor networks. *IEEE Communications Surveys & Tutorials*, 16(2):856–876, 2013.

[177] Diego Méndez, Alfredo J. Pérez, Miguel A. Labrador, and Juan José Marrón. P-sense: A participatory sensing system for air pollution monitoring and control. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 344–347, 2011.

[178] Gustavo Olivares, Ian Longley, and Guy Coulson. Development of a low-cost device for observing indoor particle levels associated with source activities in the home. *International Society of Exposure Science (ISES), Seattle, WA*, 2012.

[179] M Udin Harun Al Rasyid, Isbat Uzzin Nadhori, Amang Sudarsono, and Yodhista Tulus Alnovinda. Pollution monitoring system using gas sensor based on wireless sensor network. *International Journal of Engineering and Technology Innovation*, 6(1):79, 2016.

[180] Tzai-Hung Wen, Joe-Air Jiang, Chih-Hong Sun, Jehn-Yih Juang, and Tzu-Shiang Lin. Monitoring street-level spatial-temporal variations of carbon monoxide in urban settings using a wireless sensor network (wsn) framework. *International journal of environmental research and public health*, 10(12):6380–6396, 2013.

[181] Nurzaman Ahmed, Hafizur Rahman, and Md I Hussain. A comparison of 802.11 ah and 802.15. 4 for iot. *Ict Express*, 2(3):100–102, 2016.

[182] Shu-Chiung Hu, You-Chiun Wang, Chiuan-Yu Huang, and Yu-Chee Tseng. Measuring air quality in city areas by vehicular wireless sensor networks. *Journal of Systems and Software*, 84(11):2005–2012, 2011.

[183] Kai-Juan Wong, Chee-Cheng Chua, and Qingyun Li. Environmental monitoring using wireless vehicular sensor networks. In *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*, pages 1–4, 2009.

[184] Giuseppe Lo Re, Daniele Peri, and Salvatore Davide Vassallo. Urban air quality monitoring using vehicular sensor networks. *Advances onto the Internet of Things: How Ontologies Make the Internet of Things Meaningful*, pages 311–323, 2014.

[185] Bluetooth SIG. Bluetooth Specifications. https://www.bluetooth.com/specifications/.

[186] Kuor-Hsin Chang. Bluetooth: a viable solution for iot? [industry perspectives]. *IEEE Wireless Communications*, 21(6):6–7, 2014.

[187] Bluetooth SIG. Understanding Bluetooth Range. https://www.bluetooth.com/learn-about-bluetooth/key-attributes/range/.

[188] D. Suriano, M. Prato, V. Pfister, G. Cassano, G. Camporeale, S. Dipinto, and M. Penza. 15 - stationary and mobile low-cost gas sensor-systems for air quality monitoring applications. In *Proceedings*. COST Association, Avenue Louise 149, 1050 Brussels, Belgium, 2015.

[189] David Hasenfratz, Olga Saukh, Silvan Sturzenegger, Lothar Thiele, et al. Participatory air pollution monitoring using smartphones. *Mobile Sensing*, 1:1–5, 2012.

[190] OASIS Open. MQTT Version 3.1.1. http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html.

[191] LoP y - Pycom -Four Network Development Board for IoT. https://docs.pycom.io/datasheets/development/lopy4/.

[192] Sensirion. Mechanical Design and Assembly Guidelines for SEN5x. https://sensirion.com/media/documents/546FBC5B/61E9586E/Sensirion_Mechanical_Design_and_Assembly_Guidelines_SEN5x.pdf.

[193] The Things Network. The Things Network. https://www.thethingsnetwork.org.

[194] Amir Nagah Elghonaimy. Lorawan for air quality monitoring system. Master's thesis, Politecnico di Torino, 2021.

[195] Gabriele Telesca. Study and development of a participative air pollution monitoring system. Master's thesis, Politecnico di Torino, 2021.

[196] CRAN. Nowcast algorithm. https://cran.r-project.org/web/packages/PWFSLSmoke/vignettes/NowCast.html.

[197] Alessandro Ricciuto. Reliable and efficient solutions for a participative air pollution monitoring system. Master's thesis, Politecnico di Torino, 2021.

[198] REPUBBLICA ITALIANA. LEGGE n. 156, 9 novembre 2021. https://www.assoporti.it/media/9874/legge-9-novembre-2021-n-156.pdf.

[199] Naouras Latiri. Exploiting lorawan for iot air pollution devices in urban areas. Master's thesis, Politecnico di Torino, 2022.

[200] Arjan Avbentem. Airtime calculator for LoRaWAN. https://avbentem.github.io/airtime-calculator/ttn/eu868/20, 9 2020.

[201] Norhane BENKAHLA, Hajer TOUNSI, Ye-Qiong SONG, and Mounir FRIKHA. Enhanced adr for lorawan networks with mobility. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 1–6, 2019.

[202] Gerboles M, Spinelle L, and Borowiak A. European union tech report: Measuring air pollution with low-cost sensors. thoughts on the quality of data measured by sensors. Technical report, European Commission, 2017.

[203] Mauro Fadda, Matteo Anedda, Roberto Girau, Giovanni Pau, and Daniele D Giusto. A social internet of things smart city solution for traffic and pollution monitoring in cagliari. *IEEE Internet of Things Journal*, 2022.

[204] Matteo Anedda, Mauro Fadda, Roberto Girau, Giovanni Pau, and Daniele Giusto. A social smart city for public and private mobility: A real case study. *Computer Networks*, 220:109464, 2023.

[205] Davide Aguiari, Giovanni Delnevo, Lorenzo Monti, Vittorio Ghini, Silvia Mirri, Paola Salomoni, Giovanni Pau, Marcus Im, Rita Tse, Mongkol Ekpanyapong, et al. Canarin ii: Designing a smart e-bike eco-system. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–6. IEEE, 2018.

[206] Davide Aguiari, Chiara Contoli, Giovanni Delnevo, and Lorenzo Monti. Smart mobility and sensing: Case studies based on a bike information gathering architecture. In *Smart Objects and Technologies for Social Good: Third International Conference, GOODTECHS 2017, Pisa, Italy, November 29-30, 2017, Proceedings 3*, pages 112–121. Springer, 2018.

[207] Rita Tse, Davide Aguiari, Lorenzo Monti, Giovanni Pau, Catia Prandi, and Paola Salomoni. On assessing the accuracy of air pollution models exploiting a strategic sensors deployment. In *ACM International Conference Proceeding Series*, pages 55–58. Association for Computing Machinery, nov 2018.

[208] Thomas Baldi, Giovanni Delnevo, Roberto Girau, and Silvia Mirri. On the prediction of air quality within vehicles using outdoor air pollution: sensors and machine learning algorithms. In *Proceedings of the ACM SIGCOMM Workshop on Networked Sensing Systems for a Sustainable Society*, pages 14–19, 2022.

[209] Luigi Russi, Paolo Guidorzi, Beatrice Pulvirenti, Giovanni Semprini, Davide Aguiari, and Giovanni Pau. Air quality and comfort characterisation within an electric vehicle cabin. In *2021 IEEE International Workshop on Metrology for Automotive (MetroAutomotive)*, pages 169–174. IEEE, 2021.

[210] Luigi Russi, Paolo Guidorzi, Beatrice Pulvirenti, Davide Aguiari, Giovanni Pau, and Giovanni Semprini. Air quality and comfort characterisation within an electric vehicle cabin in heating and cooling operations. *Sensors*, 22(2):543, 2022.

[211] Rita Tse, Marcus Im, Su-Kit Tang, Luís Menezes, Alfredo Dias, and Giovanni Pau. Self-adaptive Sensing IoT Platform for Conserving Historic Buildings and Collections in Museums. In *Proceedings of the 5th International Conference on Internet of Things, Big Data and Security*, pages 392–398. SCITEPRESS - Science and Technology Publications, may 2020.

[212] Rita Tse, Davide Aguiari, Ka-Seng Chou, Su-Kit Tang, Daniele Giusto, and Giovanni Pau. Monitoring cultural heritage buildings via low-cost edge computing/sensing platforms: The biblioteca joanina de coimbra case study. In *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good*, pages 148–152, 2018.

[213] Giovanni Pettorru, Mauro Fadda, Roberto Girau, Matteo Anedda, and Daniele Giusto. An iot-based electronic sniffing for forest fire detection. In *2023 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–5. IEEE, 2023.

[214] Andrew A. Cook, Göksel Mısırlı, and Zhong Fan. Anomaly Detection for IoT Time-Series Data: A Survey. *IEEE Internet of Things Journal*, 7(7):6481–6494, July 2020.

[215] Ling-Jyh Chen, Yao-Hua Ho, Hsin-Hung Hsieh, Shih-Ting Huang, Hu-Cheng Lee, and Sachit Mahajan. ADF: An Anomaly Detection Framework for Large-Scale PM2.5 Sensing Systems. *IEEE Internet of Things Journal*, 5(2):559–570, April 2018. Conference Name: IEEE Internet of Things Journal.

[216] Erika Brattich, Alessandro Bracci, Alessandro Zappi, Pietro Morozzi, Silvana Di Sabatino, Federico Porcù, Francesca Di Nicola, and Laura Tositti. How to Get the Best from Low-Cost Particulate Matter Sensors: Guidelines and Practical Recommendations. *Sensors*, 20(11):3073, January 2020.

[217] David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin, Martin Fierz, Tabita Arn, Jan Beutel, and Lothar Thiele. Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing*, 16:268–285, January 2015.

[218] Brian Rumburg, Richard Alldredge, and Candis Claiborn. Statistical distributions of particulate matter and the error associated with sampling frequency. *Atmospheric Environment*, 35:2907–2920, June 2001.

[219] Leigh R. Crilley, Marvin Shaw, Ryan Pound, Louisa J. Kramer, Robin Price, Stuart Young, Alastair C. Lewis, and Francis D. Pope. Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmospheric Measurement Techniques*, 11(2):709–720, February 2018.

[220] Daniel Nicklin and Hamidreza Gohari Darabkhani. Techniques to measure particulate matter emissions from stationary sources: a critical technology review using multi criteria decision analysis (MCDA). *Journal of Environmental Management*, 296:113167, 2021.

[221] David H. Hagan and Jesse H. Kroll. Assessing the accuracy of low-cost optical particle sensors using a physics-based approach. *Atmospheric Measurement Techniques*, 13(11):6343–6355, November 2020.

[222] Inversione termica — Arpa Piemonte. https://www2.arpa.piemonte.it/approfondimenti/glossario/inversione-termica.

[223] N-DIMENSIONAL CUMULATIVE FUNCTION, AND OTHER USEFUL FACTS ABOUT GAUSSIANS AND NORMAL DENSITIES.

[224] Rafael Páez, Manuel Pérez, Gustavo Ramírez, Juan Montes, and Lucas Bouvarel. An architecture for biometric electronic identification document system based on blockchain. *Future Internet*, 12(1):10, 2020.

[225] Aji Supriyanto and Khabib Mustofa. E-gov readiness assessment to determine E-government maturity phase. *Proceeding - 2016 2nd International Conference on Science in Information Technology, ICSITech 2016: Information Science for Green Society and Environment*, pages 270–275, 2017.

[226] Registraduria nacional del estado civil. https://www.registraduria.gov.co/. Accessed: 2019-09-18.

[227] A. Q. Ansari. E - Document retrieval using rough-set theory. *ICIIP 2011 - Proceedings: 2011 International Conference on Image Information Processing*, 2011.

[228] Ulrich Waldmann, Sven Vow, T Sven, and Andreas Poller. Electronic Identity Cards for User Authentication – Promise and Practice. *IEEE Security and Privacy*, pages 1–15, 2012.

[229] Policia nacional de colombia. https://www.policia.gov.co/. Accessed: 2019-09-18.

[230] Md Asraful Haque, Namra Zia Khan, and Gulnar Khatoon. Authentication through keystrokes: What you type and how you type. *Proceedings of 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN*, pages 257–261, 2016.

[231] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *White Paper*, pages 1–9, 2013.

[232] Konstantinos Christidis and Michael Devetsikiotis. Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, 4:2292–2303, 2016.

[233] Thomas Bocek, Bruno B. Rodrigues, Tim Strasser, and Burkhard Stiller. Blockchains everywhere - A use-case of blockchains in the pharma supply-chain. *Proceedings of the IM 2017 - 2017 IFIP/IEEE International Symposium on Integrated Network and Service Management*, pages 772–777, 2017.

[234] Nguyen Giang Truong and Kim Kyungbaek. A survey about consensus algorithms used in Blockchain. *Journal of Information Processing Systems*, 14(1):101–128, 2018.

[235] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. An overview of blockchain technology: Architecture, consensus, and future trends. In *2017 IEEE International Congress on Big Data (BigData Congress)*, pages 557–564, 2017.

[236] Christopher Natoli and Vincent Gramoli. The balance attack or why forkable blockchains are ill-suited for consortium. In *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 579–590. IEEE, 2017.

[237] Siamak Solat and Maria Potop-Butucaru. Zeroblock: Preventing selfish mining in bitcoin. *arXiv preprint arXiv:1605.02435*, 2016.

[238] Ittay Eyal and Emin Gün Sirer. Majority is not enough: Bitcoin mining is vulnerable. *Communications of the ACM*, 61(7):95–102, 2018.

[239] A. Singh, T.-W. Ngan, P. Druschel, and D. S. Wallach. Eclipse attacks on overlay networks: Threats and defenses. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pages 1–12, 2006.

[240] Ethan Heilman, Alison Kendler, Aviv Zohar, and Sharon Goldberg. Eclipse attacks on bitcoin's peer-to-peer network. In *24th {USENIX} Security Symposium ({USENIX} Security 15)*, pages 129–144, 2015.

[241] Xiangfu Zhao, Zhongyu Chen, Xin Chen, Yanxia Wang, and Changbing Tang. The dao attack paradoxes in propositional logic. In *2017 4th International Conference on Systems and Informatics (ICSAI)*, pages 1743–1746, 2017.

[242] Aggelos Kiayias and Giorgos Panagiotakos. On trees, chains and fast transactions in the blockchain. In *Progress in Cryptology–LATINCRYPT 2017: 5th International Conference on Cryptology and Information Security in Latin America, Havana, Cuba, September 20–22, 2017, Revised Selected Papers 5*, pages 327–351. Springer, 2019.

[243] Maria Apostolaki, Aviv Zohar, and Laurent Vanbever. Hijacking bitcoin: Routing attacks on cryptocurrencies. In *2017 IEEE symposium on security and privacy (SP)*, pages 375–392. IEEE, 2017.

[244] Roberto Saia, Salvatore Carta, Diego Reforgiato Recupero, Gianni Fenu, et al. Internet of entities (ioe): A blockchain-based distributed paradigm for data exchange between wireless-based devices. In *SENSORNETS 2019-Proceedings of the 8th International Conference on Sensor Networks*, pages 77–84. SciTePress, 2019.

[245] Zhiqing Huang, Xiongye Su, Yanxin Zhang, Changxue Shi, Hanchen Zhang, and Luyang Xie. A decentralized solution for iot data trusted exchange based-on blockchain. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pages 1180–1184. IEEE, 2017.

[246] Gunnar Prause. E-residency: a business platform for industry 4.0? *Entrepreneurship and Sustainability Issues*, 3(3):216, 2016.

[247] Oscar Delgado-Mohatar, Julian Fierrez, Ruben Tolosana, and Ruben Vera-Rodriguez. Blockchain and biometrics: A first look into opportunities and challenges. In *Blockchain and Applications: International Congress*, pages 169–177. Springer, 2020.

[248] Paco Garcia. Biometrics on the blockchain. *Biometric Technology Today*, 2018(5):5–7, 2018.

[249] Ministerio del Interior España. Guía De Refernecia Del DNIE Con NFC. Technical report, Ministerio del Interior España, 2015.

[250] D. Hanchez. *A Comparative Study of Software Protection Tools Suited for E-Commerce with Contributions to Software Watermarking and Smart Cards*. PhD thesis, Univesité Chatolique de Louvain, 2003.

[251] Mindaugas Andrulevičius. Methods and applications of optical holography. *Material Science*, 17(4):371–377, 2011.

[252] Hao Tang Chan, Wen Jyi Hwang, and Chau Jern Cheng. Digital hologram authentication using a hadamard-based reversible fragile watermarking algorithm. *IEEE/OSA Journal of Display Technology*, 11(2):193–203, 2015.

[253] Bruce Schneier. One-way hash functions. In *Applied Cryptography, Second Edition*, pages 429–459. John Wiley & Sons, Inc., October 2015.

[254] Richard Duncan. An Overview of Different Authentication Methods Author Retains Full Rights tu , A ho ll r igh. *White Paper: SANS Institute*, 2001.

[255] R H Nagel. System and method for production and authentication of original documents. https://www.google.com/patents/US7080041, 2006.

[256] Bichlien Hoang and Ashley Caudill. Biometrics. Technical report, IEEE, 2012.

[257] Simon Liu and Mark Silverman. Practical guide to biometric security technology. *IT Professional*, 3(1):27–32, 2001.

[258] U.S National Science and Technology Council. Biometrics in Goverment POST-9/11. Technical report, U.S National Science and Technology Council, 2008.

[259] A.K. Jain, Jianjiang Feng, and Karthik Nandakumar. Fingerprint matching. *Computer*, 43(2):36–44, 2010.

[260] Sim Hiew Moi, Nazeema Binti Abdul Rahim, Puteh Saad, Pang Li Sim, Zalmiyah Zakaria, and Subariah Ibrahim. Iris biometric cryptography for identity document. In *SoCPaR 2009 - Soft Computing and Pattern Recognition*, pages 736–741, 2009.

[261] James Wayman, Anil Jain, Davide Maltoni, and Dario Maio. *Biometric Systems*. Springer Science & Business Media, 2005.

[262] Siddhesh Ashok Vaidya and Varsha Bhosale. Invisible touch screen based PIN authentication to prevent shoulder surfing. In *2016 International Conference on Inventive Computation Technologies (ICICT)*, pages 1–4, 2016.

[263] Vrizlynn L L Thing and Hwei-ming Ying. Rainbow Table Optimization for Password Recovery. *International Journal on Advances in Software*, 4(3):479–488, 2011.

[264] Francisco Martín-Fernández, Pino Caballero-Gil, and Cándido Caballero-Gil. Authentication Based on Non-Interactive Zero-Knowledge Proofs for the Internet of Things. *Sensors*, 16(1), 2016.

[265] Roy Fisher, Mengxuan Lyu, Bo Cheng, and Gerhard Hancke. Public key cryptography: Feasible for security in modern personal area sensor networks? *Proceedings of the IEEE International Conference on Industrial Technology*, 2016-May:2020–2025, 2016.

[266] Stephen Murrell and Norman G. Einspruch. Electronic identification, personal privacy and security in the services sector. *5th International Conference Service Systems and Service Management - Exploring Service Dynamics with Science and Innovative Technology, ICSSSM'08*, 2008.

[267] Yongsheng Liu, Jie Li, and Mohsen Guizani. PKC based broadcast authentication using signature amortization for WSNs. *IEEE Transactions on Wireless Communications*, 11(6):2106–2115, 2012.

[268] Kenneth Otula Sigar and Omari Kebiro Jared. A critical look of ussd technology adoption and benefits. *International Journal of Advanced Research in Computer Science*, 5(1):27–29, 2014.

[269] Zhu Siyang. Deformed Two-Dimension Code Quick Recognition Algorithm Design and Implementation in Uncertain Environment. *Proceedings - 2015 7th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2015*, pages 322–325, 2015.

[270] Manuel Leithner, Peter Kieseberg, Sebastian Schrittwieser, Lindsay Munroe, Martin Mulazzani, Mayank Sinha, and Edgar Weippl. QR code security. *MoMM'2010 - The Eighth International Conference on Advances in Mobile Computing and Multimedia*, page 430, 2010.

[271] Wira Firdaus Hj Yaakob, Hafizul Hasni Manab, and Siti Noorashikin Md Adzmi. Smart card chip design implementation on ARM processor-based FPGA. *2014 IEEE 3rd Global Conference on Consumer Electronics, GCCE*, pages 294–297, 2014.

[272] C. M. Roberts. Radio frequency identification (RFID). *Computers and Security*, 25(1):18–26, 2006.

[273] Sanjay E Sarma, Stephen A Weis, and Daniel W Engels. RFID Systems and Security and Privacy Implications. *Cryptographic Hardware and Embedded Systems*, pages 454–469, 2003.

[274] M.A. Zamora A.J. Jara, A.F. Alcolea and A.F.G. Skarmeta. Analysis of different techniques to define metadata structure in NFC/RFID cards to reduce access latency, optimize capacity, and guarantee integrity. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 10(Part 1):192–197, 2010.

[275] Nahar Sunny Suresh Shobha, Kajarekar Sunit Pravin Aruna, Manjrekar Devesh Parag Bhagyashree, and Kotian Siddhanth Jagdish Sarita. NFC and NFC payments: A review. *Proceedings of 2016 International Conference on ICT in Business, Industry, and Government, ICTBIG 2016*, pages 1–7, 2016.

[276] Montes D. Juan, Rincón P. Andrés, Páez M. Rafael, Ramírez E. Gustavo, and Pérez C. Manuel. A Model for National Electronic Identity Document and Authentication Mechanism Based on Blockchain. *International Journal of Modeling and Optimization*, 8(3):160–165, 2018.

[277] Emin Gün Sirer Ittay Eyal. How to disincentivize large bitcoin mining pools. http://hackingdistributed.com/2014/06/18/how-to-disincentivize-large-bitcoin-mining-pools/. Accessed: 2019-09-18.

[278] Andrew Miller, Ahmed Kosba, Jonathan Katz, and Elaine Shi. Nonoutsourceable scratch-off puzzles to discourage bitcoin mining coalitions. *Proceedings of the ACM Conference on Computer and Communications Security*, October:680–691, 2015.

[279] Scott Nadal Sunny King. PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16*, 1919(January):1–27, 2017.

[280] Pavel Vasin. BlackCoin's Proof-of-Stake Protocol v2 Pavel. *White paper*, page 2, 2014.

[281] Iddo Bentov, Ariel Gabizon, and Alex Mizrahi. Cryptocurrencies without proof of work. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9604 LNCS(240258):142–157, 2016.

[282] Mitar Milutinovic, Warren He, Howard Wu, and Maxinder Kanwal. Proof of Luck: An efficient blockchain consensus protocol. *SysTEX 2016 - 1st Workshop on System Software for Trusted Execution, colocated with ACM/IFIP/USENIX Middleware 2016*, pages 2–7, 2016.

[283] Hong Bao and David Roubaud. Non-fungible token: A systematic review and research agenda. *Journal of Risk and Financial Management*, 15(5), 2022.

[284] Lennart Ante. Smart contracts on the blockchain–a bibliometric analysis and review. *Telematics and Informatics*, 57:101519, 2021.

[285] Mingli Wu, Kun Wang, Xiaoqin Cai, Song Guo, Minyi Guo, and Chunming Rong. A comprehensive survey of blockchain: From theory to iot applications and beyond. *IEEE Internet of Things Journal*, 6(5):8114–8154, 2019.

[286] Hossein Shafagh, Lukas Burkhalter, Anwar Hithnawi, and Simon Duquennoy. Towards blockchain-based auditable storage and sharing of iot data. In *Proceedings of the 2017 on cloud computing security workshop*, pages 45–50, 2017.

[287] Mays Alshaikhli, Tarek Elfouly, Omar Elharrouss, Amr Mohamed, and Najmath Ottakath. Evolution of internet of things from blockchain to iota: A survey. *IEEE Access*, 10:844–866, 2021.

[288] Abu Buker Siddique, Rafaqat Kazmi, Habib Ullah Khan, Sikandar Ali, Ali Samad, and Gulraiz Javaid. An intelligent and secure air quality monitoring system using neural network algorithm and blockchain. *IETE Journal of Research*, pages 1–14, 2022.

[289] Cameron Thouati de Tazoult, Raja Chiky, and Valentin Foltescu. A distributed pollution monitoring system: The application of blockchain to air quality monitoring. In *Computational Collective Intelligence: 11th International Conference, ICCCI 2019, Hendaye, France, September 4–6, 2019, Proceedings, Part II 11*, pages 688–697. Springer, 2019.

[290] Shajulin Benedict, P Rumaise, and Jaspreet Kaur. Iot blockchain solution for air quality monitoring in smartcities. In *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pages 1–6. IEEE, 2019.

[291] Yohan Han, Byungjun Park, and Jongpil Jeong. A novel architecture of air pollution measurement platform using 5g and blockchain for industrial iot applications. *Procedia Computer Science*, 155:728–733, 2019.

[292] Daniele Sofia, Nicoletta Lotrecchiano, Paolo Trucillo, Aristide Giuliano, and Luigi Terrone. Novel air pollution measurement system based on ethereum blockchain. *Journal of Sensor and Actuator Networks*, 9(4):49, 2020.

[293] Quynh H. Dang. Secure hash standard. Technical report, National Institute of Standards and Technology, July 2015.

[294] Renate Nummela Caine and Geoffrey Caine. *Making Connections: Teaching and the Human Brain*. Association for Supervision and Curriculum Development, 11141 Georgia Avenue, Suite 200, Wheaton, MD 20902 (ASCD Stock No, 1991.

[295] Evan F. Risko, Nicola Anderson, Amara Sarwal, Megan Engelhardt, and Alan Kingstone. Everyday Attention: Variation in Mind Wandering and Memory in a Lecture. *Applied Cognitive Psychology*, 26(2):234–242, 2012. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.1814.

[296] Evan F. Risko, Dawn Buchanan, Srdan Medimorec, and Alan Kingstone. Everyday attention: Mind wandering and computer use during lectures. *Computers & Education*, 68:275–283, 2013.

[297] B. A. Campbell, H. Hayne, R. Richardson, and Byron A. Campbell. *Attention and information Processing in infants and Adults: Perspectives From Human and Animal Research*. Psychology Press, 2014. Google-Books-ID: oJF7AgAAQBAJ.

[298] K. Wilson and J.H. Korn. Attention during Lectures: Beyond Ten Minutes. *Teaching of Psychology*, 34(2):85–89, 2007.

[299] Ali Darvishi, Hassan Khosravi, Shazia Sadiq, and Barbara Weber. Neurophysiological Measurements in Higher Education: A Systematic Literature Review. *International Journal of Artificial Intelligence in Education*, 2021.

[300] Jiahui Xu and Baichang Zhong. Review on portable EEG technology in educational research. *Computers in Human Behavior*, 81:340–349, 2018.

[301] Genaro Rebolledo-Mendez, Ian Dunwell, Erika A. Martínez-Mirón, María Dolores Vargas-Cerdán, Sara de Freitas, Fotis Liarokapis, and Alma R. García-Gaona. Assessing NeuroSky's Usability to Detect Attention Levels in an Assessment Exercise. In Julie A. Jacko, editor, *Human-Computer Interaction. New Trends*, Lecture Notes in Computer Science, pages 149–158, Berlin, Heidelberg, 2009. Springer.

[302] Paul Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17 2:124–9, 1971.

[303] Paul Ekman. *Basic Emotions*, chapter 3, pages 45–60. John Wiley & Sons, Ltd, 1999.

[304] Daniel T. Cordaro, Rui Sun, Dacher Keltner, Shanmukh Vasant Kamble, Niranjan Huddar, and Galen D McNeil. Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion*, 18:75–93, 2018.

[305] Paul Ekman and Wallace V. Friesen. *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.

[306] Maja Pantic and Léon Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91:1370–1390, 2003.

[307] Y.-I. Tian, T. Kanade, and J.F. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115, 2001. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[308] Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro. Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1113–1133, June 2015. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[309] I.A. Essa and A.P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[310] Kenji Mase. Recognition of Facial Expression from Optical Flow. *IEICE TRANSACTIONS on Information and Systems*, E74-D(10):3474–3483, 1991. Publisher: The Institute of Electronics, Information and Communication Engineers.

[311] Zhentao Liu, Min Wu, Weihua Cao, Luefeng Chen, Jianping Xu, Ri Zhang, Mengtian Zhou, and Junwei Mao. A facial expression emotion recognition based human-robot interaction system. *IEEE/CAA Journal of Automatica Sinica*, 4(4):668–676, 2017. Conference Name: IEEE/CAA Journal of Automatica Sinica.

[312] T. Kanade, J.F. Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 46–53, 2000.

[313] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998.

[314] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces*, ICMI '04, pages 205–211, New York, NY, USA, 2004. Association for Computing Machinery.

[315] Ian Goodfellow, Dumitru Erhan, Pierre Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, and Y. Bengio. Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, 64, 07 2013.

[316] Emad Barsoum, Cha Zhang, Cristian Canton Ferrer, and Zhengyou Zhang. Training deep networks for facial expression recognition with crowd-sourced label distribution. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI '16, page 279–283, New York, NY, USA, 2016. Association for Computing Machinery.

[317] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 2106–2112, 2011.

[318] Pedro M. Ferreira, Filipe Marques, Jaime S. Cardoso, and Ana Rebelo. Physiological Inspired Deep Neural Networks for Emotion Recognition. *IEEE Access*, 6:53930–53943, 2018. Conference Name: IEEE Access.

[319] Si Miao, Haoyu Xu, Zhenqi Han, and Yongxin Zhu. Recognizing Facial Expressions Using a Shallow Convolutional Neural Network. *IEEE Access*, 7:78000–78011, 2019. Conference Name: IEEE Access.

[320] Ping Liu, Shizhong Han, Zibo Meng, and Yan Tong. Facial Expression Recognition via a Boosted Deep Belief Network. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1805–1812, 2014. ISSN: 1063-6919.

[321] Zibo Meng, Ping Liu, Jie Cai, Shizhong Han, and Yan Tong. Identity-Aware Convolutional Neural Network for Facial Expression Recognition. In *2017*

12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017), pages 558–565, 2017.

[322] Jacopo Sini, Antonio Costantino Marceddu, and Massimo Violante. Automatic Emotion Recognition for the Calibration of Autonomous Driving Functions. *518*, 9(3), 2020.

[323] Salah Rifai, Yoshua Bengio, Aaron Courville, Pascal Vincent, and Mehdi Mirza. Disentangling Factors of Variation for Facial Expression Recognition. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012*, Lecture Notes in Computer Science, pages 808–822, Berlin, Heidelberg, 2012. Springer.

[324] Su Liu, Ye Chen, Hui Huang, Liang Xiao, and Xiaojun Hei. Towards Smart Educational Recommendations with Reinforcement Learning in Classroom. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pages 1079–1084, 2018. ISSN: 2470-6698.

[325] Hamed Monkaresi, Nigel Bosch, Rafael A. Calvo, and Sidney K. D'Mello. Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate. *IEEE Transactions on Affective Computing*, 8(1):15–28, 2017. Conference Name: IEEE Transactions on Affective Computing.

[326] Arindam Ray and Amlan Chakrabarti. Design and Implementation of Technology Enabled Affective Learning Using Fusion of Bio-physical and Facial Expression. *Journal of Educational Technology & Society*, 19(4):112–125, 2016. Publisher: International Forum of Educational Technology & Society.

[327] J. Allen and A. Murray. Effects of filtering on multisite photoplethysmography pulse waveform characteristics. In *Computers in Cardiology, 2004*, pages 485–488, 2004.

[328] John Allen. Photoplethysmography and its application in clinical physiological measurement. physiological measurement. *Physiological Measurement*, 28(3):R1–R39, Feb 2007.

[329] M. J. Drinnan, J. Allen, and A. Murray. Relation between heart rate and pulse transit time during paced respiration physiol. *Physiological measurement*, 22,3(1):425–32, aug 2001.

[330] Joanne Wai Yee Chung, Henry Chi Fuk So, Marcy Ming Tak Choi, Vincent Chun Man Yan, and Thomas Kwok Shing Wong. Artificial intelligence in education: Using heart rate variability (hrv) as a biomarker to assess emotions objectively. *Computers and Education: Artificial Intelligence*, 2:100011, 2021.

[331] Atul Sharma, Mihaela Badea, Swapnil Tiwari, and Jean Louis Marty. Wearable biosensors: An alternative and practical approach in healthcare and disease monitoring. *Molecules*, 26(3), Feb 2021.

[332] Shan Xia, Shixin Song, Fei Jia, and Guanghui Gao. A flexible, adhesive and self-healable hydrogel-based wearable strain sensor for human motion and physiological signal monitoring. *Journal of Materials Chemistry B*, 7(1):4638–4648, Jun 2019.

[333] K. Ashton. That Internet of Things Thing. *RFID Journal*, 2009.

[334] Luigi Atzori, Roberto Girau, Salvatore Martis, Virginia Pilloni, and Marco Uras. A SIoT-aware approach to the resource management issue in mobile crowdsensing. In *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, pages 232–237, 2017. ISSN: 2472-8144.

[335] Luigi Atzori, Roberto Girau, Virginia Pilloni, and Marco Uras. Assignment of Sensing Tasks to IoT Devices: Exploitation of a Social Network of Objects. *IEEE Internet of Things Journal*, 6(2):2679–2692, 2019. Conference Name: IEEE Internet of Things Journal.

[336] Luigi Atzori, Claudia Campolo, Bin Da, Roberto Girau, Antonio Iera, Giacomo Morabito, and Salvatore Quattropani. Enhancing Identifier/Locator Splitting Through Social Internet of Things. *IEEE Internet of Things Journal*, 6(2):2974–2985, 2019. Conference Name: IEEE Internet of Things Journal.

[337] Luigi Atzori, Claudia Campolo, Bin Da, Roberto Girau, Antonio Iera, Giacomo Morabito, and Salvatore Quattropani. Smart devices in the social loops: Criteria and algorithms for the creation of the social links. *Future Generation Computer Systems*, 97:327–339, 2019.

[338] Xieling Chen, Di Zou, Haoran Xie, and Fu Lee Wang. Past, present, and future of smart learning: a topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, 18(1):2, 2021.

[339] Ping Tan, Han Wu, Peng Li, and He Xu. Teaching management system with applications of rfid and iot technology. *Education Sciences*, 8(1), 2018.

[340] Mohammad Ali and Al Maruf Hassan. Developing applications for voice enabled iot devices to improve classroom activities. In *2018 21st International Conference of Computer and Information Technology (ICCIT)*, pages 1–4, 2018.

[341] Ghazal Yadav, Prabha Sundaravadivel, and Lokeshwar Kesavan. Affect-learn: An iot-based affective learning framework for special education. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pages 1–5, 2020.

[342] Muhammad Awais, Mohsin Raza, Nishant Singh, Kiran Bashir, Umar Manzoor, Saif Ul Islam, and Joel J. P. C. Rodrigues. Lstm-based emotion detection using physiological signals: Iot framework for healthcare and distance learning in covid-19. *IEEE Internet of Things Journal*, 8(23):16863–16871, 2021.

[343] Antonio Costantino Marceddu, Jacopo Sini, Massimo Violante, and Bartolomeo Montrucchio. A novel approach to improve the social acceptance of autonomous driving vehicles by recognizing the emotions of passengers. In Wolfgang Osten, Dmitry P. Nikolaev, and Jianhong Zhou, editors, *Thirteenth International Conference on Machine Vision*, volume 11605, pages 503 – 510. International Society for Optics and Photonics, SPIE, 2021.

[344] François Chollet et al. Keras. https://keras.io, 2015.

[345] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 94–101, 2010.

[346] Natalie Ebner, Michaela Riediger, and Ulman Lindenberger. Faces—a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research methods*, 42:351–62, 02 2010.

[347] Niki Aifanti, Christos Papachristou, and Anastasios Delopoulos. The mug facial expression database. In *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*, pages 1–4, 2010.

[348] Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel H. J. Wigboldus, Skyler T. Hawk, and Ad van Knippenberg. Presentation and validation of the radboud faces database. *Cognition and Emotion*, 24(8):1377–1388, 2010.

[349] Jacopo Sini, Antonio Costantino Marceddu, Massimo Violante, and Riccardo Dessì. Passengers' emotions recognition to improve social acceptance of autonomous driving vehicles. In Anna Esposito, Marcos Faundez-Zanuy, Francesco Carlo Morabito, and Eros Pasero, editors, *Progresses in Artificial Intelligence and Neural Systems*, pages 25–32, Singapore, 2021. Springer Singapore.

[350] Antonio Costantino Marceddu, Jacopo Sini, Massimo Violante, and Bartolomeo+ Montrucchio. A novel approach to improve the social acceptance of autonomous driving vehicles by recognizing the emotions of passengers. In *Proceedings of SPIE*, Rome, 2021. SPIE.

[351] Antonio Costantino Marceddu. Emotion detector (ed). https://github.com/AntonioMarceddu/Emotion_Detector, 2020.

[352] Eclipse Deeplearning4j Development Team. Deeplearning4j: Open-source distributed deep learning for the jvm, apache software foundation license 2.0. https://deeplearning4j.konduit.ai/.

[353] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference*

on Computer Vision and Pattern Recognition. CVPR 2001, volume 1, pages I–I, 2001.

[354] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 21–37, Cham, 2016. Springer International Publishing.

[355] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[356] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[357] James Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 12 1980.

[358] Steve Walsh. *Exploring classroom discourse: Language in action*. Routledge, 2011.

[359] Jacopo Sini. Reaction time tool. https://github.com/JacopoSini/MDPI-Sensors---IOT-Education-ReactionTimeTool, 2021.