Monitoring Outdoor Air Quality Using Personal Device to Protect Vulnerable People

Luca Rampini, Fulvio Re Cecconi
Department of Architecture, Built Environment and Construction Engineering
Politecnico di Milano
Milan, Italy
e-mail: {luca.rampini, fulvio.rececconi}@polimi.it

***Abstract*—World Health Organization considered air pollution the most dangerous threat to human health. This paper presents a novel system to mitigate risks derived from polluted air by introducing a portable device for vulnerable people able to detect hazardous pollution concentration and to suggest healthy behaviors. Along with the device, a tailored web server application, called Monitoring Outdoor Quality of Air (MOQA), is developed. The application will make the data collected from the device's sensors more readable and will provide citizens with air quality information. That information will alert vulnerable people about air pollution hazards and let them take actions consequently. This system can reduce the health risks derived from air pollution for the weakest population in the short-term.**

Keywords-Outdoor Air Quality; Pollution Sensor; CO_2 Sensor; Machine Learning; Sensor reliability.

I. INTRODUCTION

SLow Onset Disasters (SLODs) are continuous, low intensity, and high frequency events that represent a serious risk for the health of people [1][2]. Among SLODs, the World Health Organization (WHO) declared air pollution as the biggest environmental threat [3]. Hence, developing a system that helps the whole community to mitigate the risks arising from polluted air would be of great benefit to everyone.

Air pollution is caused by a range of factors including transportation, agriculture, waste, anthropogenic activities (e.g., respiration), construction, and building operations. Indeed, 39% of global energy-related carbon emissions are related to the building sector [4]. Within this context, Built Environment scenarios, that have low capacity to deal with heat absorption and rejection are more prone to expose their hosts (people) to SLODs.

Moreover, the impact of air pollution on people is well documented, revealing that some demographic categories might be more disadvantaged than others. This vulnerable population comprehends toddlers [5], elders [6], and those suffering from respiratory issues (e.g., allergies, asthma) or cardiovascular problems [7]. Considering that 1) the arousal frequency of SLODs' evidence is larger in urbanized context [8], 2) more than half of the world's population lives in urban areas [9], handling strategies should be a priority.

The effects of interactions between humans and their Built Environment (BE) have been studied for a long time [10]-[12]. At the same time, it is well documented the influence of BE in affecting the intensity of SLODs [13]-[16]. Consequently, this trine relationship among people, BE and SLOD, represents a complex mechanism to depict.

In Europe, researchers have been studied the progression of SLODs, identifying where these disasters are more prone to take place or to have a larger effect. For instance, the work in [3] reported Italy as one of the most critical areas of Central Europe, where the worst conditions have been shown for the municipality of Milan. Therefore, this territory represents a useful case study to test procedures that might reduce air pollution effects.

Currently, air pollution can be noted and predicted but is harder to mitigate. This study aims to reduce risks derived from polluted air, by introducing a system that might help to protect vulnerable people. Exploiting the novel field of Citizen Science [17], in which ordinary people contribute information for scientific research, a wearable device has been made. The device can measure data about CO_2, Particulate Matter (PM), and Total Volatile Organic Compounds (TVOCs), and send them to a tailored web server application, called Monitoring Outdoor Quality of Air (MOQA). Looking to the application, vulnerable people might understand when the quality conditions of the air are dangerous for their health and consequently, take precautions. In a short-term period, the actions taken by alerted people can mitigate the risks deriving from the most extreme dangers, reducing the incidence of pollution on the weaker population [18].

This study is unique in the following ways: i) to the best of our knowledge, this is the first research that comprehends the introduction of tailored software and a personal portable device that can retrieve air quality data; ii) the involvement of others citizen to collect air quality measurements, exploiting the principles of Citizen Science; iii) the implementation of an alert system to increase the safety of vulnerable people by mitigating the risks derived from air pollutants concentration event.

The rest of this paper is organized as follows: Section II presents the background of this research. In Section III are reported methods and tools, that comprehend hardware and software description. Those instruments will be deployed in a case study described in Section IV. Finally, the conclusion and future works are discussed in Section V.

Copyright (c) IARIA, 2020.

II. Background

Typically, air quality is monitored by stationary monitors, able to measure pollutants as well as temperature and relative humidity. However, these monitoring stations are difficult to operate, bulky, and expensive. Each instrument can cost from about 6000 euros to tens of thousands euros [19], in addition to a relevant amount of resources to be dedicated to maintenance. Moreover, those monitors are sparsely distributed in specific locations, therefore, they might represent pollutant levels, that vary a lot spatially and temporally [20], only in limited areas near the stations. Hence, recently, there has been an emerging trend of using low-cost, portable sensors to provide air quality data as an alternative monitoring option. Indeed, compared to the more expensive air monitors, the introduced personal sensors had an overall cost of less than one hundred euros. Those sensors are easy to deploy, operate, and maintain, thus present significant characteristic to be used at multiple locations [21].

However, low-cost sensors generate data with lower precision and accuracy than those registered by monitors [22], thus several studies develop calibration models deploying Machine Learning (ML) techniques to improve the performance of the low-cost sensors. For instance, Alhasa et al., used a Multi-Layer Perceptron (MLP) to calibrate three EC AlphaSense sensors, obtaining an r score in the range of 0.89 to 0.92 [23]. Among different ML algorithms, Neural Networks (NNs) and ensemble methods (e.g., XGBoost) produce the best results, with the former performing generally slightly better than the latter, albeit requires greater efforts in tuning the hyperparameters [24].

III. Methods and Tools

The method adopted in this study is represented in Figure 1, reporting the Research Schema. The first step involves volunteer who collects data as a part of a scientific inquiry, adopting the principles of Citizen Science. This information is useful to depict the effects given by the interactions between humans and the built environment, which can influence SLODs trend. The measurements are collected exploiting the usability of the mobile software application and low-cost sensors devices. The portable device, unlike the stationary monitors, could provide representative data for a whole urban area, helping to form a clearer picture of pollution field trends.

To facilitate involved citizens' work, a tailored mobile software application is introduced to retrieve and store the data measured by the portable device. To calibrate and validate those data, a measurement campaign was carried out near the air quality monitor sensors situated in Città Studi. The data gathered with this system are then compared with those collected by the ARPA (Agenzia Regionale per la Protezione Ambientale) station.

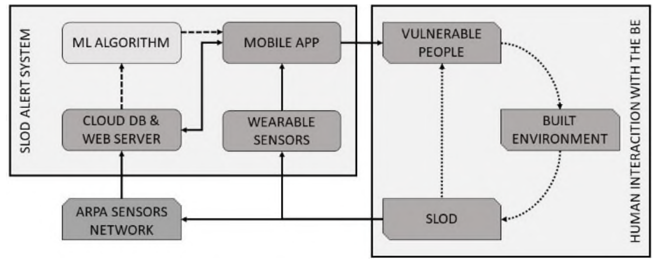

Figure 1. Research Schema.

The calibration is carried out by an Artificial Neural Network (ANN) that takes the data from low-cost sensors as input and transform them into corrected measurements using ARPA dataset.

The ANN has allowed increasing the precision of the device's measurements, reducing the difference with the pollutant level measured by the station. Once the model is trained, citizens can measure air pollution in different areas using the data corrected by the ANN and, in case of a dangerous situation, modify their behavior (e.g., bring their inhaler or decide to remain at home). Finally, in the next sub-sections are described the hardware components used to build the portable device, the developed software application, and the data taken from the regional open data repository.

A. Hardware

To accomplish the objectives of the Research Schema, this study introduced a portable device able to measure pollutant levels in the air as well as atmospheric information such as temperature and relative humidity. The main board of the device is an Arduino, whereas the following sensors measure information about air quality:

- 1 SparkFun Environmental Combo Breakout - CCS811/BME280 - SEN-14348: this sensor measures Temperature, Relative Humidity (RH), Atmospheric pressure as well as pollutants' amount such as equivalent CO_2 (eCO_2) and Total Volatile Organic Compound (TVOC). The eCO_2 output range is from 400 to 8192 ppm, whereas the TVOC output range is from 0 to 1187 ppb.
- 1 PM sensor SDS011: this sensor detects PM in the air. The PM output range is from 0 to 999 $\mu g/m^3$ and the humidity working range of the sensor is 0-70%.

The core part of the wearable device is shown in Figure 2.

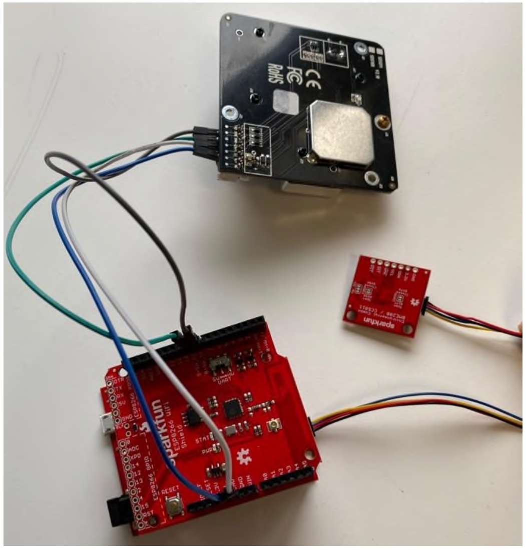

Figure 2. Core sensors of the portable device.

The device makes a measurement every 90 seconds and sends it to the web application via wi-fi.

B. Software

To collect, store, and visualize data measured by the device, this research has introduced a new cross-platform application called Monitoring Outdoor Quality of Air (MOQA).

To fetch data and send them to the server, the users should be registered via email and password. The database used is a Relational DataBase Management System (DBMS), which guaranteed the fundamental properties for a database of this type: i) atomicity, ii) consistency, iii) isolation and, iv) durability. The entity-relationship diagram (Figure 3) shows the main sources of the application.

To encourage widespread adoption of the app, a Graphical User Interface (GUI) has been developed and the features of the application are divided by screens as follows:

- Authentication: the user log-in the application by his/her email and password;
- Arduino: the user has a quick view of the last data retrieved by the wearable device. Besides, it is also possible to send and store the data to the remote server;
- Map (Figure 4): this screen displays the data coming from the Arduino board (blue circles) and ARPA dataset (red circles) on a map. The radius of a circle depends on the value of the data associated;
- Chart: the user can see the data coming from the Arduino board and ARPA dataset on a line plot. Moreover, a table with a computation of the quartile deviation is provided;
- Filter (Figure 5): The user could pick the measure to visualize. Once the category is selected, the user can choose the time range and whether to show the ARPA data corresponding to a defined station;
- Settings: the user can change his/her personal information, as well as edit the IP address and port where Arduino is available.

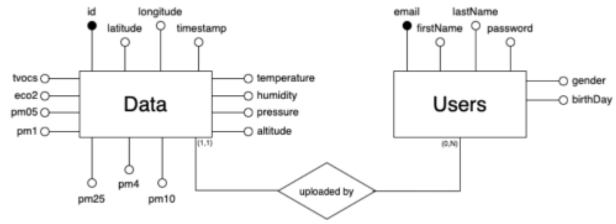

Figure 3. The entity-relationship diagram of the database.

Figure 4. MOQA map screen.

In Figure 6, it is shown the use case diagram, taken from the documentation available on the GitHub repository of the MOQA application [25], that shows the flow of operations triggered by an actor, which could be the user or the system that wants to perform a task. Inside the diagram, the actions colored in light blue are performed by the application whereas the green one represents the data source.

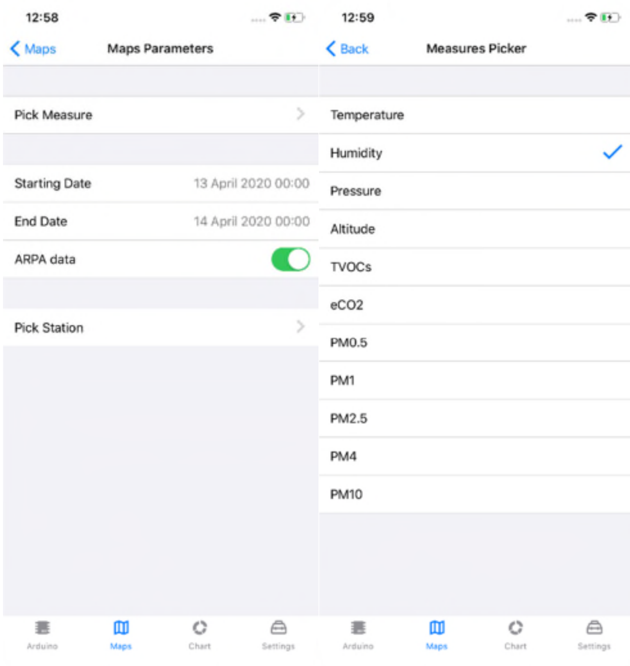

Figure 5. Filter screens.

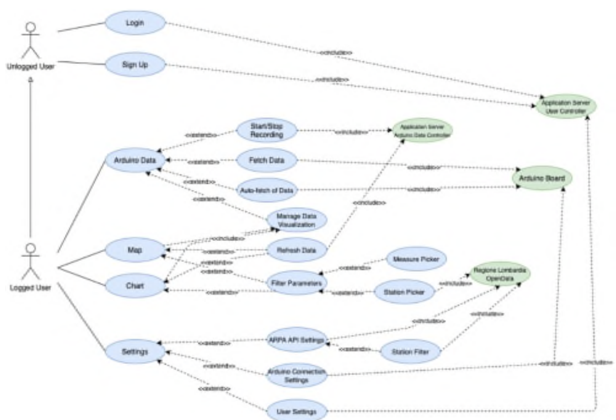

Figure 6. Use case diagram of the system [25].

The major benefits driven by MOQA will be its usability and its data reading easiness that allows non-expert users to monitor air quality conditions.

C. ARPA Data

The data collected with the portable device are compared to those downloaded from the Lombardy Region's open data repository (ARPA Lombardia). This database is quite extensive and contains all the measures registered by the stations, divided into yearly historical series. Each yearly dataset for each Air Quality (AQ) station is composed of approximately 2.5 million records. Therefore, data for all the available AQ sensors have been cleaned, removing negative or absent values, and keeping only the validated (VA) records. The analysis of the data provided by ARPA can provide useful insights about how the pollutants are distributed throughout the year and what affects them. For instance, in Figure 7, it is reported the analysis of PM10 data distribution from 2017 to February 2020. It is clear how the PM10 concentration is dependent on the heating period in Italy for climate zone E: 15th October to 15th April as stated by the DPR n. 74/2013. Indeed, in those periods, the total amount of PM10 exceeds the threshold of 50 μg/m3 set by the European Environment Agency [26].

IV. RESULTS

The portable sensor had been tested against a reference one, i.e., the monitor situated in Città Studi and managed by ARPA, to verify its reliability. More than five hundred measures of CO2, temperature, and relative humidity of the two sensors included in the portable device were used in the test.

Figure 8 shows a comparison between the simultaneous reading of the portable sensor and a reference sensor. The difference between the measure of the reference sensor and the portable one is computed according to the following formula:

$$\Delta CO_2 = {(CO_{2,r} - CO_{2,p})}/{CO_{2,r}} \qquad (1)$$

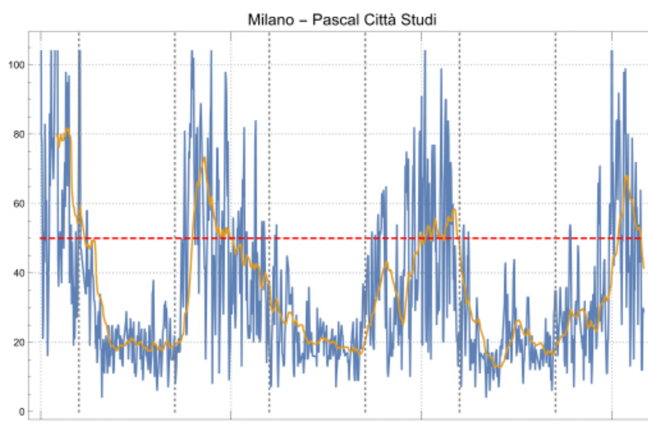

Figure 7. PM10 from 01/01/2017 to February 2020 registered by the AQ station Pascal Città Studi (ARPA). The blue line represents the daily average, whereas the orange line shows the 1 month moving average.

Where:
ΔCO_2 is the difference between two measures
$CO_{2,r}$ is the reference sensor measure
$CO_{2,p}$ is the portable sensor measure

The reliability of the sensor is not suitable for most scientific applications, since during the test period 73.64% of the ΔCO_2 were either bigger than 50% or lower than -50%.

The difference between the reference and the portable sensor readings may depend on many factors. Among these, temperature and relative humidity can influence the data collected by the device. Figure 9, it shows the difference between the two sensors CO2 measures as a function of the difference between temperature and relative humidity readings. For instance, in some temperature conditions,

Copyright (c) IARIA, 2020. ISBN: 978-1-61208-820-4

Figure 9 highlights a direct correlation between relative humidity and CO2 error.

To calibrate the device's sensor to the monitor's measures a deep learning model is implemented. The ANN designed is a feedforward neural network [27] with an input layer, three hidden layers and, an output layer. The net uses two types of hidden layer: a) linear layer, i.e., a trainable, fully connected net layer that computes w·x+b, where w represents the weights, x the input features, and b the bias; b) Tanh layer, i.e., a layer that applies the hyperbolic tangent function to its input. The net has twelve neurons on each layer except for the input and the last two layers. To train the network, 80% of the available data were fed to it, whereas the remaining 20% has been used for test evaluation.

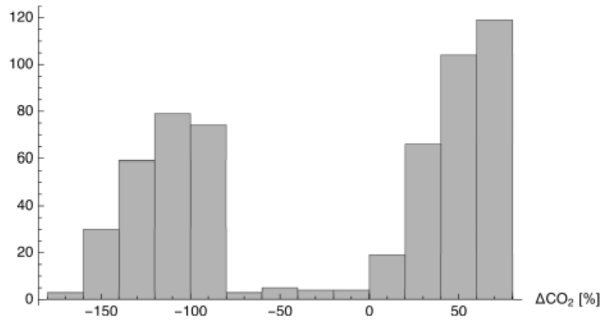

Figure 8. The percentage difference between reference and portable sensor readings.

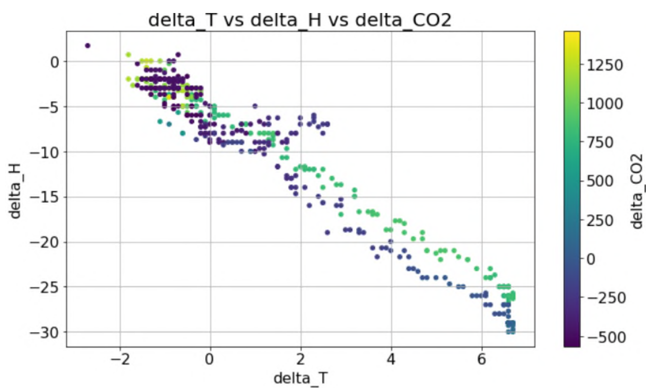

Figure 9. Differences in CO_2 sensor readings (color scale) as a function of the difference of temperature (X axes) and relative humidity (Y axes).

The performance of the network on the test dataset is shown in the scatter plot in Figure 10. The trained net performs very well since the Pearson's correlation coefficient computed for actual and forecasted CO2 is R2=97.59.

Summing up, the ANN allows for much better precision. As reported in Figure 11, the difference between portable sensor's measures improved using the ANN and the ones of the reference sensor computed according to (1). The ΔCO_2, in this case, is between -5% and +5% in 92.97% of the readings, hence the errors that occurred with non-calibrated data are almost canceled. Moreover, most of the incorrect correlations are around 900 ppm values (both predicted and actual), thus a further investigation on that aspect might hone the results. In conclusion, the data adjusted by the network might be used to make measures in different zones far from the reference station and thus, help to depict polluted areas dangerous for vulnerable people.

V. Conclusion and Future Work

Although SLODs are among the major causes of death, few studies have been carried out on their diffusion in the built environment, especially in big cities. There are too few air quality monitoring stations in cities to understand the exact relationship between pollutants and the built environment. A crowdsourcing data collection on SLODs may help to solve this issue, but low-cost sensor devices are needed.

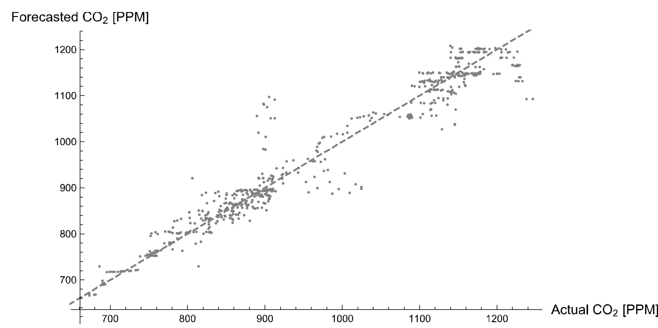

Figure 10. Scatter plot of the ANN performance.

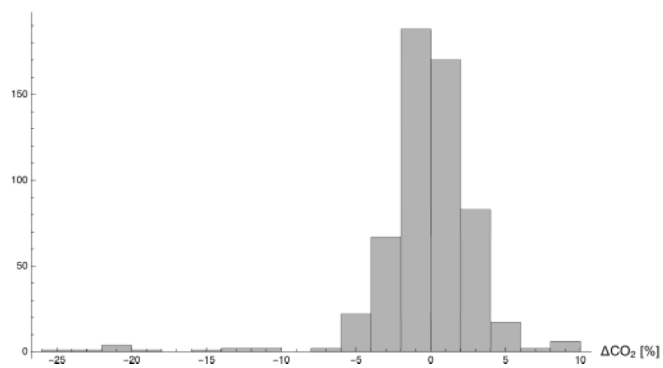

Figure 11. The percentage difference between reference and portable sensor readings after ANN correction.

Thus, a portable air quality monitoring device had been made. The small and cheap device and the bespoke mobile application allowed for citizen involvement in collecting data, but their reliability appeared to be too low during a test with reference sensors. To overcome this hurdle, ML had been used to calibrate the portable sensor. This approach proved to be effective and the data collected from citizens are now sufficiently reliable to continue with the next stages of research.

Future works will include the implementation of a warning system inside the mobile application that will alert vulnerable people in case of dangerous conditions and make them act accordingly. This step will help non-expert users to be aware of the risks without the need to have a thematic knowledge of air pollution and pollutants threshold levels.

Meanwhile, the work on the portable device will continue and will be twofold: i) optimization of the hardware (better connection, a 3D printed shell, on board software optimization, etc) and ii) try to improve machine learning performance by deploying further approaches with higher model complexity.

ACKNOWLEDGMENT

The authors would like to thank Mathays Giudici for his valuable work in the development of the mobile application.

FUNDING

This research was partially funded by the MIUR (the Italian Ministry of Education, University, and Research) Project BE S2ECURe - (make) Built Environment Safer in Slow and Emergency Conditions through behavioral assessed/designed Resilient solutions (Grant number: 2017LR75XK).

REFERENCES

[1] J. E. Jackson et al., "Public health impacts of climate change in Washington State: Projected mortality risks due to heat events and air pollution," Clim. Change, vol. 102, no. 1–2, pp. 159–186, 2010, doi: 10.1007/s10584-010-9852-3.

[2] J. Y. Lee and H. Kim, "Projection of future temperature-related mortality due to climate and demographic changes," Environ. Int., vol. 94, pp. 489–494, 2016, doi: 10.1016/j.envint.2016.06.007.

[3] World Health Organization, "Ambient Air Pollution: A global assessment of exposure and burden of disease," 2016.

[4] WorldGBC, "About Green Building," World Green Building Councils, 2020. [Online]. Available: https://www.worldgbc.org/what-green-building.

[5] P. de Prado Bert, E. M. H. Mercader, J. Pujol, J. Sunyer, and M. Mortamais, "The Effects of Air Pollution on the Brain: a Review of Studies Interfacing Environmental Epidemiology and Neuroimaging," Current environmental health reports, vol. 5, no. 3. pp. 351–364, 2018, doi: 10.1007/s40572-018-0209-9.

[6] R. B. Devlin, A. J. Ghio, H. Kehrl, G. Sanders, and W. Cascio, "Elderly humans exposed to concentrated air pollution particles have decreased heart rate variability," in European Respiratory Journal, Supplement, 2003, vol. 21, no. 40, doi: 10.1183/09031936.03.00402403.

[7] R. J. Delfino et al., "Traffic-related air pollution and blood pressure in elderly subjects with coronary artery disease," Epidemiology, vol. 21, no. 3, pp. 396–404, 2010, doi: 10.1097/EDE.0b013e3181d5e19b.

[8] EPA, "Urban heat island basics," in Reducing Urban Heat Islands: Compendium of Strategies | Heat Island Effect | US EPA, 2008, p. 22.

[9] World Urbanization Prospects: The 2018 Revision. 2019.

[10] N. Garin et al., "Built Environment and Elderly Population Health: A Comprehensive Literature Review," Clin. Pract. Epidemiol. Ment. Heal., vol. 10, no. 1, pp. 103–115, 2014, doi: 10.2174/1745017901410010103.

[11] A. Forsyth, J. Michael Oakes, B. Lee, and K. H. Schmitz, "The built environment, walking, and physical activity: Is the environment more important to some people than others?," Transp. Res. Part D Transp. Environ., 2009, doi: 10.1016/j.trd.2008.10.003.

[12] S. Srinivasan, L. R. O'Fallon, and A. Dearry, "Creating Healthy Communities, Healthy Homes, Healthy People: Initiating a Research Agenda on the Built Environment and Public Health," American Journal of Public Health, vol. 93, no. 9. pp. 1446–1450, 2003, doi: 10.2105/AJPH.93.9.1446.

[13] I. D. Stewart and T. R. Oke, "Local climate zones for urban temperature studies," Bull. Am. Meteorol. Soc., vol. 93, no. 12, pp. 1879–1900, 2012, doi: 10.1175/BAMS-D-11-00019.1.

[14] R. Paolini, A. G. Mainini, T. Poli, and L. Vercesi, "Assessment of thermal stress in a street canyon in pedestrian area with or without canopy shading," in Energy Procedia, 2014, doi: 10.1016/j.egypro.2014.02.177.

[15] E. Jamei, P. Rajagopalan, M. Seyedmahmoudian, and Y. Jamei, "Review on the impact of urban geometry and pedestrian level greening on outdoor thermal comfort," Renewable and Sustainable Energy Reviews. 2016, doi: 10.1016/j.rser.2015.10.104.

[16] N. Colaninno and E. Morello, "Modelling the impact of green solutions upon the urban heat island phenomenon by means of satellite data," in Journal of Physics: Conference Series, 2019, doi: 10.1088/1742-6596/1343/1/012010.

[17] R. Lukyanenko, A. Wiggins, and H. K. Rosser, "Citizen Science: An Information Quality Research Frontier," Inf. Syst. Front., vol. 22, no. 4, pp. 961–983, 2020, doi: 10.1007/s10796-019-09915-z.

[18] H. Chen et al., "Effect of air quality alerts on human health: a regression discontinuity analysis in Toronto, Canada," Lancet Planet. Heal., vol. 2, no. 1, pp. e2–e3, 2018, doi: 10.1016/S2542-5196(17)30185-7.

[19] A. Cavaliere et al., "Development of low-cost air quality stations for next generation monitoring networks: Calibration and validation of PM2.5 and PM10 sensors," Sensors (Switzerland), vol. 18, no. 9, 2018, doi: 10.3390/s18092843.

[20] Y. Xiong, J. Zhou, J. J. Schauer, W. Yu, and Y. Hu, "Seasonal and spatial differences in source contributions to PM2.5 in Wuhan, China," Sci. Total Environ., vol. 577, pp. 155–165, 2017, doi: 10.1016/j.scitotenv.2016.10.150.

[21] D. M. Holstius, A. Pillarisetti, K. R. Smith, and E. Seto, "Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California," Atmos. Meas. Tech., vol. 7, no. 4, pp. 1121–1131, 2014, doi: 10.5194/amt-7-1121-2014.

[22] R. Jayaratne, X. Liu, P. Thai, M. Dunbabin, and L. Morawska, "The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog," Atmos. Meas. Tech., vol. 11, no. 8, pp. 4883–4890, 2018, doi: 10.5194/amt-11-4883-2018.

[23] K. M. Alhasa et al., "Calibration model of a low-cost air quality sensor using an adaptive neuro-fuzzy inference system," Sensors (Switzerland), vol. 18, no. 12, 2018, doi: 10.3390/s18124380.

[24] M. Si, Y. Xiong, S. Du, and K. Du, "Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods," Atmos. Meas. Tech., vol. 13, no. 4, pp. 1693–1707, 2020, doi: 10.5194/amt-13-1693-2020.

[25] Mathyas Giudici, "MOQA design and technology." [Online]. Available: https://github.com/MathyasGiudici/polimi-dima-moqa-documentation/blob/master/dima-moqa-documentation.pdf.

[26] R. Derwent and A.-G. Hjellbrekke, Air quality in Europe — 2019 report 1994-2019 EEA Report No 10/2019. 2019.

[27] G. Bebis and M. Georgiopoulos, "Why network size is so important," IEEE Potentials, vol. 13, no. 4, pp. 27–31, 1994, doi: 10.1109/45.329294.