Framework and Method for Measurement of Particulate Matter Concentration using Low Cost Sensors

Shree Vidya Gurudath*[1]
Ramaiah Institute of Technology
MSR Nagar, Bangalore, India

Krishna Raj P M[2]
Ramaiah Institute of Technology
MSR Nagar, Bangalore, India

Srinivasa K G[3]
NITTTR
Chandigarh, India

***Abstract*—Rapid urbanisation and infrastructure shortcomings leading to heavy traffic, heavy construction activities are major contributors to emission of particulate matter into the ambient atmosphere. This is especially true in developing countries, such as India and China. There have been numerous attempts from government authorities and civic agencies to curtail pollution, but these efforts have been in vain. Cities like Beijing, New Delhi suffer from extremely unhealthy air quality during multiple months of the year. Hence, the onus of keeping oneself safe from extreme affects of air pollution falls on the individual. The following study presents a method and framework to measure particulate matter (PM2.5) concentration using low cost sensors, and infer patterns from the data collected. The study uses a SDS011 high precision laser PM2.5 detector module combined with a raspberry pi, which communicates the measurements through *message queueing telemetry transport* (MQTT) protocol to a ponte server which inturn persists the data into a MongoDB, which can be consumed by algorithms for further analysis. For example, the data obtained from the sensors can be fused with data from measurement stations and geographical land use information to estimate dense spatio-temporal pollution maps which is the basis for computing individual exposure to pollutants.**

Keywords*—Air pollution; low cost sensor; optical dust sensors; particulate matter; MQTT; ponte

I. Introduction

Particulate matter is a major source of air pollution across the world. Exposure to particulate matter can cause a multitude of problems to individuals including higher risk of hypertension Prabhakaran et al. [1], lung cancer Ciabattini et al. [2], cardio-vascular disease Jaafari et al. [3]. The affects of pollution are more prominent in the most vulnerable subset of population such as pregnant women Tapia et al., Zhu et al. [4, 5], infants Zhou et al., Rivera et al. [6, 7], and the elderly Wang et al., Han et al. [8, 9].

Recently, there have been major strides in quantifying the exposure of individuals to pollutants. Government agencies report air quality in the form of *Air Quality Index* (AQI), which signifies short term effects of pollutants in the atmosphere. Additionally, there have been numerous studies exploring the use of low cost and mobile sensors to measure individual exposure in an efficient manner. Karagulian et al. [10] provides a comprehensive review of various studies for measurement of different pollutants using low cost sensors.

There have been a few studies where the sensors are mounted on a moving vehicle, and pollution mapping is done along the path of the vehicle. DeSouza et al. [11] mounted Alphasense OPC-N2 sensors on garbage trucks to map out air quality in Cambridge, MA. The study used the collected data to identify clusters signifying pollutant hotspots, and explored techniques to generalise the measurements across the entire traversed route.

Low cost sensors are an efficient way for measuring individual exposure to pollutants. Mahajan and Kumar [12] explored the usage of low cost sensors to quantify individual exposure. The study used a PMS5003 sensor which published data using an ESP8255 wifi module. Similar to this study, calibration of the PMS5003 sensor was performed by the colocation technique with a GRIMM EDM 107 dust monitor. Motlagh et al. [13] provide a vision for the use of low cost sensors for dense air quality monitoring and the study also documents an example implementation in the city of Helsinki, Finland.

Chen et al. [14] used a SDL-607 to measure school students' exposure to PM2.5. The sensor measures particulate matter concentration by means of laser scattering using the principle of nephelometry. The data was stored in internal memory and manually transferred to a computer. Candia et al. [15] propose a system and framework for using low cost sensor networks for air quality monitoring. The study used Nova SDS011, SDS021, and SHINYEI PPD42 sensors connected to an Arduino Mega, with an ESP8266 wifi module and a LoRaWAN module in the absence of a wifi network.

Most of these studies concentrate on the efficacy of the sensors, but there are not many studies which propose a comprehensive framework for acquisition and management of pollution data effectively. There are some platforms available which have been used, for example Schneider et al. [16] used *AQMesh* (AQMesh [17]) platform to map out urban pollution in Oslo, Norway. AQMesh provides a proprietary solution for data management which includes web, API and other modes of access. Lim et al. [18] used *AirCasting* (HabitatMap [19]), an opensource data visualisation platform, to map out pollution in Seoul, South Korea.

In the following sections, a detailed description of the proposed framework is provided. Firstly, the sensors utilized and their internal components are described. Section II-B describes the calibration process for the sensors to avoid drift and inaccuracies in the reading. Finally, Section II-C describes the overall data acquisition process, including the edge devices, protocol and server infrastructure involved.

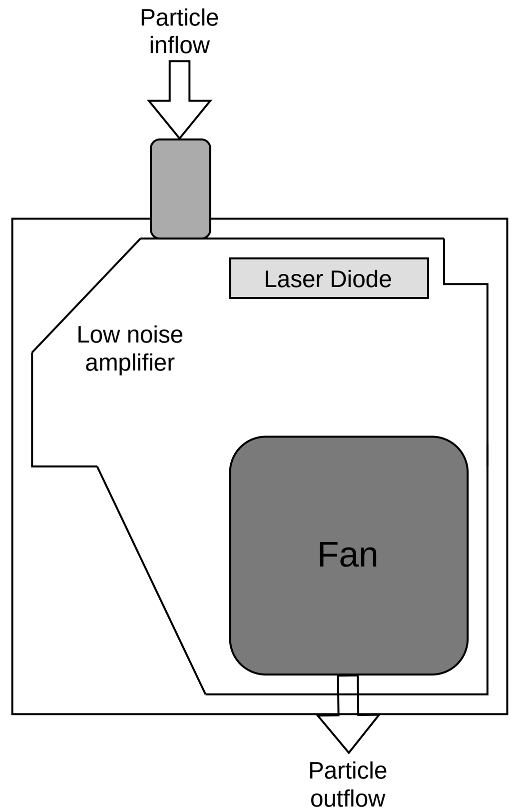
II. Measurement Framework

A. Sensors

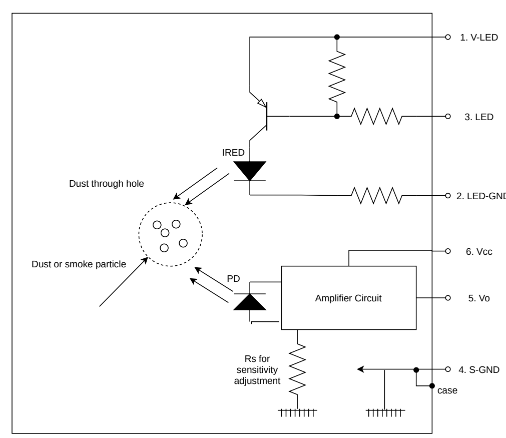
During the study, two sensors were evaluated viz., a Sharp GP2Y1010AU0F optical sensor and a SDS011 laser sensor. Both the GP2Y1010AU0F and the SDS011 sensors use light scattering principle to count the number of particles in the air sample. While the Sharp sensor uses infra-red light, the SDS011 sensor uses laser scattering. Fig. 4 shows the circuit diagram used to connect the sensor to an Arduino Uno in order to get the particulate matter concentration. The SDS011 sensor is easier to use. It has a UART connector that can be directly connected to raspberry Pi using a USB connection. Another advantage of the SDS011 sensor is that it has an inbuilt fan, in order to ensure uniform distribution of the particles in the measurement air sample.

SDS011 is a dust sensor developed by Nova fitness, which uses laser diffraction principle to measure the concentration of particulate matter in the air sample. 1a shows a schematic diagram of the internal components of the SDS011 sensor. It has a sensor output frequency of 1Hz, with a sensitivity range of $0 - 999.9\mu gm^{-3}$. The sensor has an inbuilt fan and is the most accurate for the size with a relative error of 10%. The sensor has a working range in 0-70% humidity and $-20 - 50^0C$.

The SHARP GP2Y1010AU0F is an optical dust sensor similar to the SDS011. b shows the inner circuit diagram of the sensor (Fig. 1). It uses an IR LED and phototransistor to measure the amount of dust in the air sample. One of the salient features of the sensor is that it can detect dust from a single pulse, hence works at much higher frequency of upto 100Hz. The sensor measurements are read as voltage which is proportional to the density of particulate matter in the air sample. The sensitivity of the sensor ranges from $0.35 - 0.65V/(0.1\mu gm^{-3})$. The main drawback of the sensor is that it does not have an inbuilt fan, hence it needs to be housed in a mixing chamber with a fan in order to get reliable readings. However, when mounted on a moving vehicle, proper placement can eliminate the need for a mixing chamber.



(a) SDS011 Sensor



(b) GP2Y1010AU0F Sensor

Fig. 1. The SDS011 and GP2Y1010AU0F Sensors.

B. Sensor Calibration

Low cost sensors are very capable of detecting pollutant concentrations effectively. With proper maintenance and calibration, they are suitable for long term usage too. Liu et al. [20] studied the performance of a Plantower PMS1003 sensor for PM2.5 measurement at two locations in Australia and China, over a period of 13 months. Zusman et al. [21] provide a comprehensive guide to calibrating low cost sensors using multiple techniques. In this study, sensor calibration was performed by co-location method.

The calibration for the SHARP GP2Y1010AU0F sensor was done by colocation technique. The sensor was colocated near Jayanagar air quality monitoring station maintained by Karnataka state pollution control board (KSPCB) located at 12.920984 LAT, 77.584908 LONG. Simultaneous readings were taken for a duration of two weeks. Fig. 2 shows the outputs of the two sensors. Pearson correlation between the SHARP sensor measurements vs reference sensor is 0.9121332.

C. Data Acquisition Infrastructure

The data measured from the sensors is published through MQTT to ponte server which provides a MQTT broker. The data is persisted in MongoDB, which is used for offline and realtime data analysis. Fig. 3 shows a Schematic representation

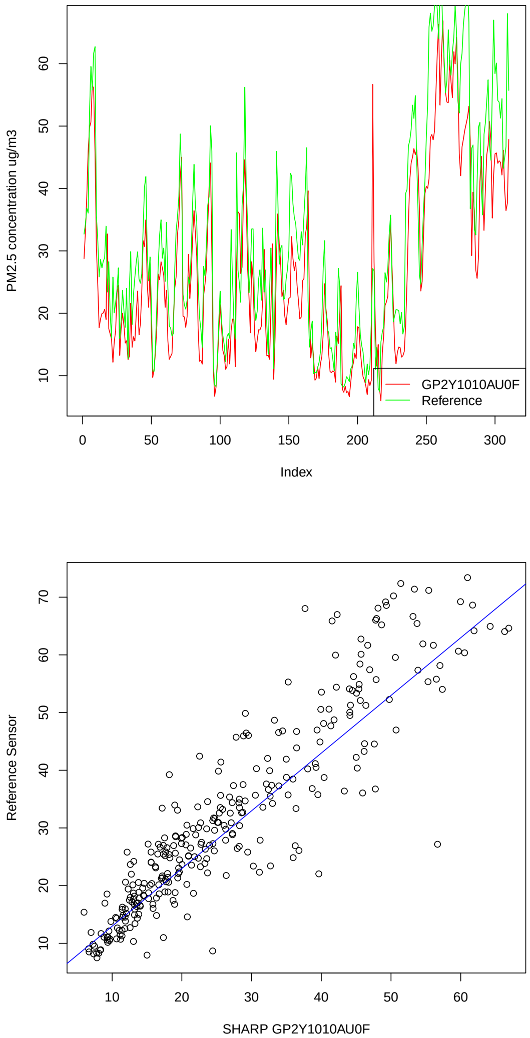


Fig. 2. SHARP Sensor Measurements VS Reference Sensor

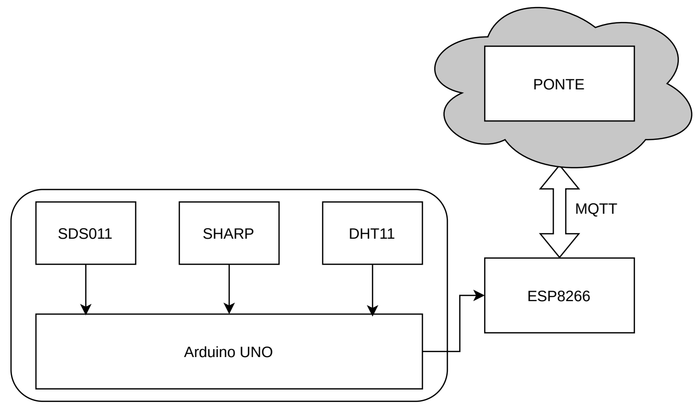


Fig. 3. Schematic Showing the Overall Data Acquisition and Persistence Framework.

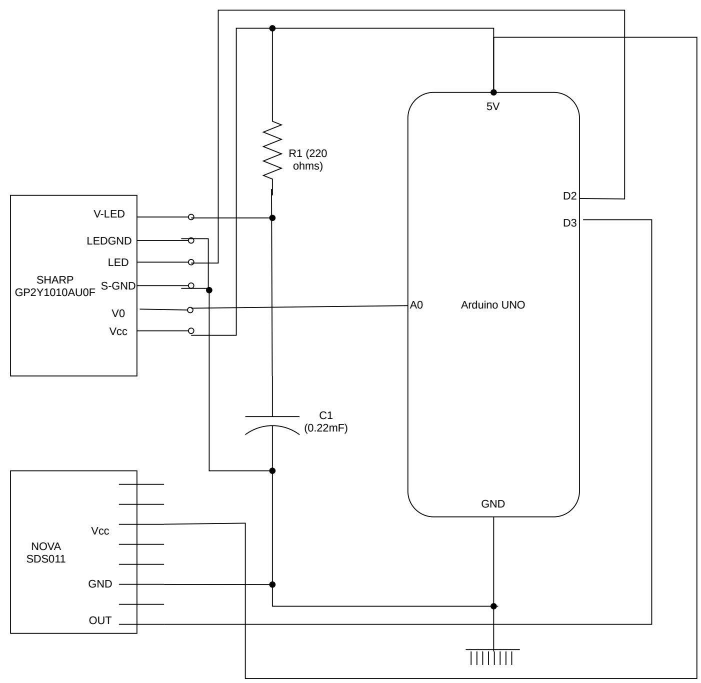


Fig. 4. Circuit Diagram to Connect the Sensors to an Arduino Uno.

of the overall data acquisition and persistence framework.

1) Message Queueing Telemetry Transport: MQTT is a publish-subscribe type messaging protocol which is widely used in IoT (Internet of Things) applications. One of the major advantages of MQTT is that the client applications are very light weight, making it ideal for deployment in small embedded systems such as low power micro-controllers. Additionally, the protocol is scalable, potentially capable of incorporating millions of devices.

Fig. 5 shows an example of an IoT system using MQTT. There are mainly three types of nodes in a MQTT network, viz., the publisher, the subscriber and the MQTT broker. A node can take the role of a subscriber or publisher or both simultaneously. The broker is part of the infrastructure and resides in a server machine. The publisher node publishes data using a specific *topic*, while subscriber nodes which have subscribed for the specific topic, will recieve the message with the data.

MQTT also provides an option for the broker to store messages when a subscriber is down, so that the messages are delivered when the client comes up. Another advantage of a publish/subscriber model over a traditional client servers is that the publisher does not need to be aware of the nature or number of subscribers. Although, the subscribers need to be aware of the message format and the topic of the publisher. MQTT and publish/subscribe model for communication is an efficient model for low power sensor networks such as the one used in the study. It also supports OAUTH based authentication, so that the data is protected from unauthorised access. It also provides for TLS/SSL data encryption.

In this study, each edge node containing an Arduino Uno is assigned a MD5 hash, which acts as an identifier string, and each node publishes the measured PM2.5 concentration, temperature and humidity using a topic which includes the identifier. For example, 84474b87cafa5f22d8aa2a5b990bfe98 is the identifier for the node which was colocated with the Jayanagar 5th Block monitoring station, and the sensor measurements from the node is published using a topic strings 84474b87cafa5f22d8aa2a5b990bfe98_dht11, 84474b87cafa5f22d8aa2a5b990bfe98_gp2y, and, 84474b87cafa5f22d8aa2a5b990bfe98_sds011 for the temperature and humidity values from the DHT11 sensors, PM2.5 concentrations from the SHARP GP2Y1010AU0F and the Nova SDS011 sensors respectively in the form of a JSON object.

2) Server Infrastructure: The data published by the edge nodes need a MQTT broker to route the messages to the subscribers. Also, there is a need for a scalable database where the data can be persisted for future analysis. Additionally, a HTTP server interface is helpful for a web based dashboard with visualization and realtime data showcasing. In this study, eclipse ponte is used, to accomplish the above requirements. Ponte provides a seamless integration with MongoDB for persisting the data as well as sessions. Additionally, it provides HTTP and CoAP services with REST like APIs.

Fig. 6 shows the schematic architecture of Ponte, with interfaces for MQTT, HTTP and CoAP (Constrained Application Protocol) protocols. The figure also shows the persistence module which connects to MongoDB which not only stores the published data, but also the session information for facilitating seamless communication even in fickle networks.

On the server, different analysis and visualisation applications can provide good insights into the individuals exposure to pollutants. Additionally, a cumulative exposure index (C) can be calculated based on pollutant concentrations exposed to and the amount of time spent. It can be computed as an integral over time, or a simplified summation (C') for discrete readings as shown in equation 1 and 2.

$$C = \int_t d_t dt \tag{1}$$

$$C' = \sum_t d_t \tag{2}$$

where d_t is the pollutant concentration at time t. The measurements, d_t can be geographically distributed across the city.

III. Discussion

Low cost sensors and their efficacy in measuring pollutant concentrations is the subject of multiple studies, hence this study does not go into the details of these factors. However, there are only a few studies which elaborate on the framework and infrastructure involved in collecting the measurements in a database and the interfaces required for analysing the data. This study provides a comprehensive guide to the processes, system and framework necessary for measurement, persistence and analysis of pollution data.

Sensors used in the current study require periodic calibration. Chain calibration technique, where a correctly calibrated low cost sensor co-located with the deployed sensors can be used to avoid sensor drift and other errors in the measurements. Additionally, temperature and humidity play a major role in the accuracy of the measurements. Hojaiji et al. [23] provide a way to calibrate the SHARP sensor to compensate the effects of temperature and humidity. The SDS011 sensor is relatively robust to temperature and humidity variations. In the current study, we smooth the variations caused due to temperature and relative humidity for the SHARP sensor. The study also ignores readings from both sensors when the humidity level rises beyond 70%. Spatial and temporal regression techniques can be used to interpolate the measurements from the monitoring stations. In a future work, a combination of long short term memory (LSTM) networks and land use regression (LUR) in order to build a dense spatio-temporal pollution map is evaluated.

MQTT is a robust transmission protocol which supports the necessary data distribution, encryption and authentication features. Publish/subscribe model of communication is also a very efficient mode for medium scale networks which is necessary for pollutant modeling for a city. However, MQTT transmit cycles are not well equipped for large networks, CoAP could be a feasible alternative when designing a network larger than 250 measurement nodes. Ponte server, used in the study, supports publishing data through CoAP using a simple REST API. This would be the focus of a future study to augment the findings presented in this article.

The framework presented above, is capable of being extendible to different types of sensors, different and more efficient protocols, more stringent and effective security measures, and different persistence technologies. The framework presented above serves as an outline for a more complex and feature rich system implementation.

A. Conclusion and Future Scope

In this paper, a system and framework for measurement, persistence and analysis of pollution data using low cost sensors was proposed. A SHARP GP2Y1010AU0F and a SDS011 dust sensors were used to measure the PM2.5 concentration in ambient atmosphere and the data was sent to a server running ponte using MQTT. The data was persisted in MongoDB which is used for further analysis on the server. Additionally, in order to compensate for the effect of temperature and relative humidity, a DHT11 sensor provided the necessary readings.

The above work is the first step towards a comprehensive estimation and prediction of individual exposure to pollutants and it's deposition in the human lungs. In the future, as a progression of this work, the authors intend to estimate dense spatio-temporal pollution maps with the fusion of data measured using these low cost sensors, and from measurement stations and provide a predictive methodology to quantify lung deposition of these pollutants for an individual.

References

[1] Dorairaj Prabhakaran, Siddhartha Mandal, Bhargav Krishna, Melina Magsumbol, Kalpana Singh, Nikhil Tandon, KM Venkat Narayan, Roopa Shivashankar, Dimple Kondal, Mohammed K Ali, et al. Exposure to particulate matter is associated with elevated blood pressure and incident hypertension in urban india. *Hypertension*, 76(4):1289–1298, 2020.
[2] Marco Ciabattini, Emanuele Rizzello, Francesca Lucaroni, Leonardo Palombi, and Paolo Boffetta. Systematic review and meta-analysis of recent high-quality studies on exposure to particulate matter and risk of lung cancer. *Environmental Research*, page 110440, 2020.

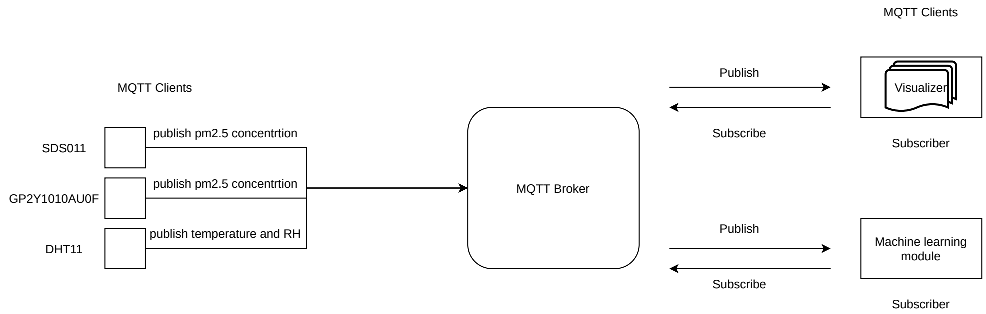


Fig. 5. MQTT Publish Subscribe Mechanism.

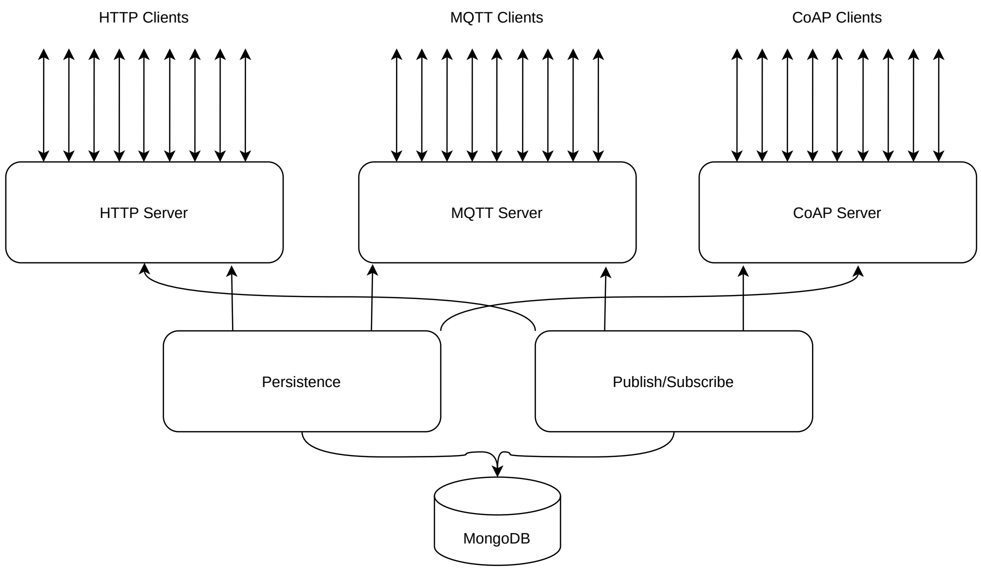


Fig. 6. Eclipse Ponte Architecture (reproduced from Eclipse [22]).

[3] Jalil Jaafari, Kazem Naddafi, Masud Yunesian, Ramin Nabizadeh, Mohammad Sadegh Hassanvand, Mansour Shamsipour, Mohammad Ghanbari Ghozikali, Hamid Reza Shamsollahi, Shahrokh Nazmara, and Kamyar Yaghmaeian. The acute effects of short term exposure to particulate matter from natural and anthropogenic sources on inflammation and coagulation markers in healthy young adults. *Science of The Total Environment*, 735:139417, 2020.

[4] VL Tapia, BV Vasquez, B Vu, Y Liu, K Steenland, and GF Gonzales. Association between maternal exposure to particulate matter (pm 2.5) and adverse pregnancy outcomes in lima, peru. *Journal of exposure science & environmental epidemiology*, 30(4):689–697, 2020.

[5] Wentao Zhu, Huiqiu Zheng, Jieyu Liu, Jiajie Cai, Gechao Wang, Yi Li, Haochong Shen, Jing Yang, Xuemei Wang, Jing Wu, et al. The correlation between chronic exposure to particulate matter and spontaneous abortion: A meta-analysis. *Chemosphere*, page 131802, 2021.

[6] Shuang Zhou, Lizi Lin, Zheng Bao, Tong Meng, Shanshan Wang, Gongbo Chen, Qin Li, Zheng Liu, Heling Bao, Na Han, et al. The association of prenatal exposure to particulate matter with infant growth: A birth cohort study in beijing, china. *Environmental Pollution*, 277: 116792, 2021.

[7] Nadya Y Rivera Rivera, Marcela Tamayo-Ortiz, Adriana Mercado García, Allan C Just, Itai Kloog, Martha Maria Téllez-Rojo, Robert O Wright, Rosalind J Wright, and Maria José Rosa. Prenatal and early life exposure to particulate matter, environmental tobacco smoke and respiratory symptoms in mexican children. *Environmental Research*, 192: 110365, 2021.

[8] Jiaonan Wang, Tiantian Li, Yuebin Lv, Virginia Byers Kraus, Yi Zhang, Chen Mao, Zhaoxue Yin, Wanying Shi, Jinhui Zhou, Tongzhang Zheng, et al. Fine particulate matter and poor cognitive function among chinese older adults: evidence from a community-based, 12-year prospective cohort study. *Environmental health perspectives*, 128(6):067013, 2020.

[9] Changwoo Han, Jongmin Oh, Youn-Hee Lim, Soontae Kim, and Yun-Chul Hong. Long-term exposure to fine particulate matter and development of chronic obstructive pulmonary disease in the elderly. *Environment International*, 143:105895, 2020.

[10] Federico Karagulian, Maurizio Barbiere, Alexander Kotsev, Laurent Spinelle, Michel Gerboles, Friedrich Lagler, Nathalie Redon, Sabine Crunaire, and Annette Borowiak. Review of the performance of low-cost sensors for air quality monitoring. *Atmosphere*, 10(9):506, 2019.

[11] Priyanka DeSouza, Amin Anjomshoaa, Fabio Duarte, Ralph Kahn, Prashant Kumar, and Carlo Ratti. Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustainable Cities and Society*, 60:102239, 2020.

[12] Sachit Mahajan and Prashant Kumar. Evaluation of low-cost sensors for quantitative personal exposure monitoring. *Sustainable Cities and Society*, 57:102076, 2020.

[13] Naser Hossein Motlagh, Eemil Lagerspetz, Petteri Nurmi, Xin Li, Samu Varjonen, Julien Mineraud, Matti Siekkinen, Andrew Rebeiro-Hargrave, Tareq Hussein, Tuukka Petaja, et al. Toward massive scale air quality monitoring. *IEEE Communications Magazine*, 58(2):54–59, 2020.

[14] L-W Antony Chen, John O Olawepo, Felicia Bonanno, Aman Gebreselassie, and Mi Zhang. Schoolchildren's exposure to pm 2.5: a student club–based air quality monitoring campaign using low-cost sensors. *Air Quality, Atmosphere & Health*, 13(5):543–551, 2020.

[15] Agustin Candia, Soledad Natacha Represa, Daniela Giuliani, Miguel Ángel Luengo, Andrés Atilio Porta, and Luis Armando Marrone. Solutions for smartcities: proposal of a monitoring system of air quality based on a lorawan network with low-cost sensors. In *2018*

Congreso Argentino de Ciencias de la Informática y Desarrollos de Investigación (CACIDI), pages 1–6. IEEE, 2018.

[16] Philipp Schneider, Nuria Castell, Matthias Vogt, Franck R Dauge, William A Lahoz, and Alena Bartonova. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environment international*, 106:234–247, 2017.

[17] AQMesh. AQMesh. https://www.aqmesh.com/products/aqmesh/, 2011. [Online; accessed 24-Oct-2021].

[18] Chris C Lim, Ho Kim, MJ Ruzmyn Vilcassim, George D Thurston, Terry Gordon, Lung-Chi Chen, Kiyoung Lee, Michael Heimbinder, and Sun-Young Kim. Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in seoul, south korea. *Environment international*, 131:105022, 2019.

[19] HabitatMap. AirCasting. https://www.habitatmap.org/aircasting, 2011. [Online; accessed 24-Oct-2021].

[20] Xiaoting Liu, Rohan Jayaratne, Phong Thai, Tara Kuhn, Isak Zing, Bryce Christensen, Riki Lamont, Matthew Dunbabin, Sicong Zhu, Jian Gao, et al. Low-cost sensors as an alternative for long-term air quality monitoring. *Environmental research*, 185:109438, 2020.

[21] Marina Zusman, Cooper S Schumacher, Amanda J Gassett, Elizabeth W Spalt, Elena Austin, Timothy V Larson, Graeme Carvlin, Edmund Seto, Joel D Kaufman, and Lianne Sheppard. Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environment international*, 134:105329, 2020.

[22] Eclipse. PONTE. https://github.com/eclipse/ponte, 2014. [Online; accessed 24-Oct-2021].

[23] Hannaneh Hojaiji, Haik Kalantarian, Alex AT Bui, Christine E King, and Majid Sarrafzadeh. Temperature and humidity calibration of a low-cost wireless dust sensor for real-time monitoring. In *2017 IEEE sensors applications symposium (SAS)*, pages 1–6. IEEE, 2017.