SENSOR NETWORK AND VISUALIZATION FOR AIR POLLUTION MONITORING

by

Ashlin Saju

THESIS DRAFT

UNIVERSITY OF NORTHERN BRITISH COLUMBIA
April 5, 2021

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Earth, our home planet, is the only known place in the universe that is affirmed to host life [14]. Our life on earth is characterized by three components, air, land, and water. Each element has its special property and is required in its proper proportions to maintain the healthy life of all living beings [14]. However, industrial development and other manmade activities create an imbalance in the natural environment. The process of making environment unsuitable and unsafe for the living condition by introducing substances that are harmful to the surroundings is called pollution [15].

Pollution changes the quality of the environment and is transboundary, as the pollutants travel thousands of miles [14]. The introduction of harmful solid, liquid, or gaseous substances from human activities into the environment changes the quality of the surroundings that we live in [15]. The presence of pollutants in the environment makes an adverse impact on human health and surrounding [16] [17]. Many types of pollutants contribute to global warming and climate change which are the major issues tackled by environmental scientists these days. Out of different types of pollution, anthropogenic or human-influenced air pollution plays a major role in climate change as well as in international public health issues [15].

Air contamination can be referred to as to the release of pollutants into the environment that is unfavorable to human wellbeing and the planet in general. According to the State of

Global Air [18] the air pollution is a complex mixture of gases and particles whose sources and composition vary over space and time.

The degrading quality of air over a region can be directly related to the development of industries, human activities, and other natural activities like volcanic eruptions, forest fires, etc. Contamination of air is a matter of serious concern and the public is often unaware of the impact that it causes to human health as well as to the surroundings. The World Health Organization (WHO) reported that the death rate estimates are around 7 million every year as 9 out of 10 people breathe polluted air [19]. This has led many motivated individuals like researchers and communities to work towards creating awareness among the people.

There has been an enormous number of studies done to understand air pollution. Most of them involved in observing one or two pollutants of interest that are dominant in the area of measurement. Other works are on improving the quality of the collected data, and also on effective visualization. In our research work, we have tried to create a complete system that measures a set of pollutants, and the data collected from the system are calibrated for improving the data quality. The calibrated data is then visualized using a software tool. In this way, we have tried to simplify the complexity of measurements for laypersons who have interests in monitoring air pollution. This approach requires having a combined background in electronics and advanced computer skills to develop such a system.

In the following section, we will discuss the background to introduce my research work and how air pollution is being currently measured in Prince George.

1.1 Background

The success of the Industrial Revolution and urbanization led to the development and growth of the economy, society, and large scale industries. This involved the use of more mechanization and the introduction of new technologies that led to the release of harmful pollutants into the environment. The introduction of a variety of pollutants into the environment created an imbalance in the ecosystem throughout the world [15].

The declining quality of air has changed from a local issue to an international public health issue. The use of coal as an energy source in industries for example in Europe and North America contributed to black smoke pollution [20]. Coal was not just used in industries but also in houses for heating in winter which made the pollution even worse [21]. These emissions resulted in serious health impacts on residents in urban areas that increased the mortality rate during the 19th century.

One such important event in the history of pollution is the great smog of London which killed as many as 12,000 people, mostly infants. This was caused due to the combination of cold weather with smoke and lasted for several days [22]. There was a string of similar events reported in New York, England, and other parts of the world around the same time. With several incidents contributing to the global pollution led to the development of various private and government entities for ensuring air quality.

Governments along with these environmental agencies established legislation like the clean air act, the motor vehicle air pollution act, air pollution control acts for a better quality of air. Apart from that, they took the initiative to monitor air pollution by installing systems that could measure the concentration of pollutants and could give warnings to the public as well as industries regarding how polluted the atmosphere is.

1.1.1 Existing Environmental Monitoring station

Government and environmental agencies are making an effort to install monitoring stations for understanding air quality. These agencies monitor the ’criteria’ pollutants (also called as common pollutants) along with any special pollutant that is dominant in that area. These monitoring stations are fixed in a location and are operated by environmental agencies. The stations are equipped with instruments that not only monitor criteria pollutants but also analyze other parameters like wind speed, humidity, precipitation. These analytical instruments work by the principle of sampling of the air collected from the atmosphere. There are two main methods for pollutant sampling: passive sampling, and active sampling [23]. These sampling techniques are considered as one of the most significant developments for air quality measurement and used widely for monitoring purposes.

In passive sampling, the pollutants are collected by a physical process such as diffusion through a static air layer or membrane. These pollutants in the air are adsorbed on the sampling media due to the chemical composition of the pollutants. The analysis of the pollutant on the sampling media gives the time-averaged contaminant concentration.

On the other hand, active sampling works with an air sampling pump which actively pulls the air through a collection device like a filter, and weighted concentration is calculated. However, these instruments have a major drawback of temporal resolution as they are large and need regular maintenance. These instruments are expensive and are financially impractical to expand to multiple stations. As an example table 1.1, gives the average estimated cost for purchasing air quality monitoring equipment produced by the US Environment Protection Agency (EPA). The FTIR in table 1.1 is the Fourier Transform Infrared spectroscopy measures multiple gases. It can be seen that cost of individual instruments for measuring the pollutants is considerable.

Pollutant/Parameter	Estimated cost
NOx	10,4440 USD
SO_2	35,000 USD
COx	28,000 USD
Ox	6,600 USD
PM	37,700 USD
FTIR Analyzer	100,000 USD

Table 1.1: Estimated cost of air quality monitoring equipment [1]

As a result of the high cost, only a few monitoring stations are installed for an area. Thus, we can claim that spatial resolution is limited to these conventional monitoring systems. This has led researchers and scientists to work on portable and less expensive sensor networks to understand air pollution in more detail.

1.2 Air Pollutants and Measurement Metrics

Various pollutants contribute to the contamination of the environment. These pollutants differ from region to region depending on human activities. For example, in an industrial area that manufactures products from raw materials, such as the production of iron from its ore or production of gasoline from crude oil, releases inorganic carbon compounds into the atmosphere [24]. These pollutants released from industrial activities can have a huge impact on human health as well as the ecosystem.

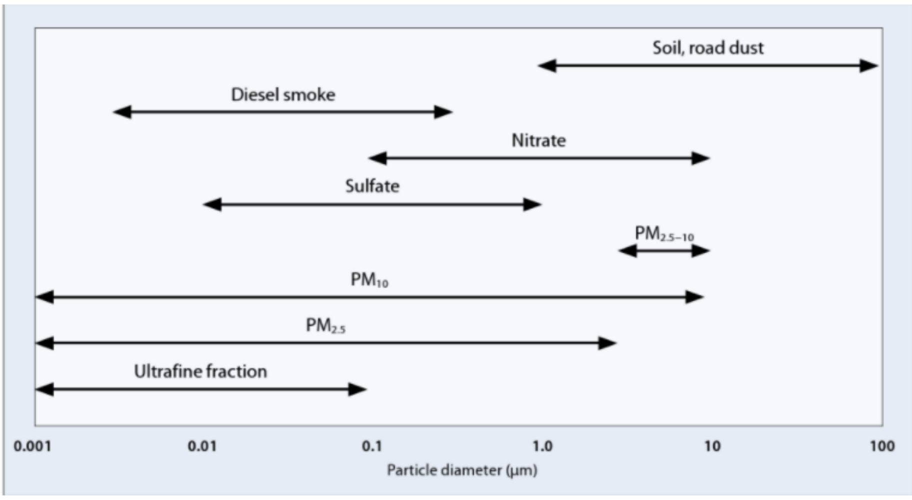


Figure 1.1: The size of Particulate Matter including PM_{10}, $PM_{2.5}$ and ultrafine particle [2]

One of the pollutants of most serious health concerns is Particulate Matter. Particulate matter is described as the particles that are formed in the atmosphere due to a chemical reaction between different pollutants in the environment [15]. These particles vary in their diameter and are measured at two levels; fine particles which are 2.5 microns or less in size ($PM_{2.5}$) and coarse particles which are 10 microns or less in size (PM_{10}). These are measured in terms of concentration in micrograms per cubic meter ($\mu g/m^3$) [25]. The fine particles are formed by foamed aerosols, metal vapors, and combustion particles. The coarse particles are formed from the break up of larger particles and contain road dust, earth crust materials, and industrial particles. The generation of both fine and coarse particles is from industrial activities, factories, construction, burning of fossil fuels, or even vacuuming. The varied composition of size can be related to the source where the particle is generated. The figure 1.1 shows the range of size that particulate matter can vary on a logarithmic scale along with the range of other components [2]. The PM_{10}, $PM_{2.5}$ along with ultrafine particles shown in figure 1.1 are the particles that cause health issues like asthma, lung disease, heart attacks, and other serious issues like cardiovascular and respiratory diseases. They are capable of penetrating through the lungs leading to cardiovascular and Chronic Obstructive Pulmonary Diseases (COPD) [7].

The next pollutant on the list is Carbon Monoxide (CO) which is produced from incomplete combustion of fossil fuels such as motor vehicle emission [26]. The majority of the CO present in the atmosphere is from road traffic and the rest is from the burning of other fuels [26]. Exposure to Carbon Monoxide which is a colorless and odorless gas, results in absorption of the gas into the bloodstream and reduces the ability of lungs to transfer oxygen which in turn affects the functionality of vital organs such as the brain and the heart [27] [28]. Another vehicle emitted gas that harms air quality is Nitrogen Dioxide (NO_2). Nitrogen is the most abundant element in earth's atmosphere and is approximately 78% by volume [29]. There are five forms of gaseous Nitrogen: Nitrogen (N_2), Nitrogen Dioxide (NO_2), ammonia (NH_3), Nitrous Oxide (N_2O), and Nitric Oxide (NO). The combustion of motor vehicle exhaust, industrial process, fuel combustion for heating produces NO which is easily oxidized to NO_2 or Nitrogen Dioxide [29].NO_2 can cause adverse pulmonary disease when inhaled in high concentrations and causes illness such as wheezing, coughing, bronchitis, and increases the severity of flu symptoms [30]. The next serious pollutant present is Ground Level Ozone (O_3) which is formed from a chemical reaction between pollutants like oxides of Nitrogen and Volatile Organic Compounds (VOCs) with sunlight [31]. Ozone can also be produced when there is a reaction between oxides of Nitrogen and hydrocarbons in the presence of sunlight [5]. The rate of production of Ozone increases at high temperature and in the presence of sunlight. Respiratory issues such as a decrease in responsiveness of airways, inflammation in airways, breathing difficulties and lung infectivity occur due to exposure of high concentration of ozone (O_3) [32]. Ozone can also decrease the productivity of vegetation and crops [33]. These are the most common pollutants seen almost everywhere but there are also other pollutants like Lead (Pb) or Benzene (C_6H_6) depending on the industrial activity in that area. All these pollutants can cause severe health impacts and also reduce life expectancy or even could cause death.

Based on the severity of health impact and the kind of human activities, different government agencies around the globe have taken measures to preserve the environment. Each country developed its indexes to identify the impact of pollution on the surroundings. For this, they might include a specific set of pollutants that are local to the region and varies from region to region.

For example in the United States, the Environmental Protection Agency (EPA) established the National Ambient Air Quality Standards (NAAQS) which specifies the pollutants that are harmful to the public health and environment. The NAAQS has a set of six common criteria pollutants that harm human health, the environment, or even cause property damage. The pollutants specified by NAAQS are Particulate Matter (PM), Ozone (O_3), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), Sulphur Dioxide (SO_2), and Lead (Pb).

India on the other hand measures eight major pollutants; Particulate matter(PM), Ozone(O_3), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), Sulphur Dioxide (SO_2), Ammonia (NH_3), and Benzene (C_6H_6) (in some places (Pb) instead). Most other countries measures a subset of these criteria pollutants, for example, Canada measures PM, O_3, NO_2, SO_2 and CO [34].

Each country has identified a set of pollutants and measurement units for these pollutants. For a layman to understand these individual measurements and its cumulative impact on the quality of air is challenging. Taking this into account, the government agencies of each country has developed their indices similar to the NAAQS for representing the quality of air. A wide range of indices have been proposed like Air Quality Health Index (AQHI), Air Quality Index (AQI), Air Pollution Index (API), Pollution Standard Index (PSI), Comprehensive Air Quality Index (CAI), Daily Air Quality Index, Common Air Quality Index

(CAQI) are few used in different countries [35]. Out of all these indexes, the most common is AQI and AQHI which are proposed and used by different countries [34].

India, USA, UK, and many other countries use AQI, and Canada, Hong Kong uses AQHI. These metrics are designed by carefully examining those pollutants which are harmful to human health and the environment. The AQI is defined as a piecewise linear function of the pollutant concentration [36] and is measured using the following formula.

$$AQI = Max\{I_i | i = 1, ..., 8\} \tag{1.1}$$

where I_i is an air quality sub-index corresponding each pollutant and it is computed as

$$I_i = \lceil(\frac{I_{high} - I_{low}}{C_{high} - C_{low}})\rceil \times (C - C_{low}) + I_{low} \tag{1.2}$$

where C is concentration of the i^{th} pollutant. C_{low} and C_{high} are lower and upper concentration breakpoints of C respectively. I_{low} and I_{high}, respectively, are index breakpoints corresponds to C_{low} and C_{high}. The value of AQI varies from 0 to 400+ as shown in Figure 1.2 and is color-coded to show the quality of air in the atmosphere. This index is described to the public by categorizing into six categories: good (0-50), satisfactory (51-100), moderately polluted (101-200), poor (201-300), very poor (301-400), and severe (>401).

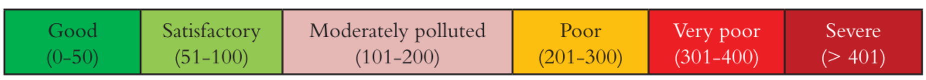


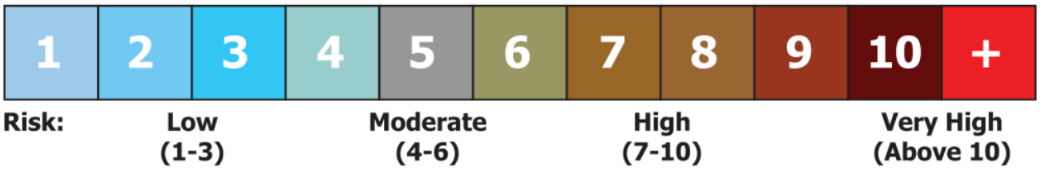
Figure 1.2: Air Quality Index (AQI) [3]

In past Canada used the USA's AQI for understanding the air quality. Due to some concerns raised by provincial and municipal health authorities, Health Canada along with Environment Canada developed AQHI to make the public aware of the quality of air that surrounds them and how it affects their health. The basis of generating this newer metrics is to understand how air quality affects human health and this is achieved by statistically linking the pollutant data with the outcomes of human mortality [37].

Initially, it was based on five major pollutants $PM_{2.5}$, O_3, NO_2, SO_2, and CO initially and later the last two pollutants were dropped from the calculation as they were identified to contribute less in predicting health effects. The following formula computes AQHI.

$$AQHI = \lceil (\frac{1000}{10.4}) \times [e^A - 1] + [e^B - 1] + [e^C - 1] \rceil \tag{1.3}$$

where $A = 0.000537 \times$ concentration of O_3 (ppb), $B = 0.000871 \times$ concentration of NO_2 (ppb) and $C = 0.000487 \times$ concentration of $PM_{2.5}$ (ug/m^3).



Health Risk	Air Quality Health Index	Health Messages: At Risk Population*	Health Messages: General Population
Low	1 - 3	**Enjoy** your usual outdoor activities.	**Ideal** air quality for outdoor activities.
Moderate	4 - 6	**Consider reducing** or rescheduling strenuous activities outdoors if you are experiencing symptoms.	**No need to modify** your usual outdoor activities unless you experience symptoms such as coughing and throat irritation.
High	7 - 10	**Reduce** or reschedule strenuous activities outdoors. Children and the elderly should also take it easy.	**Consider reducing** or rescheduling strenuous activities outdoors if you experience symptoms such as coughing and throat irritation.
Very High	Above 10	**Avoid** strenuous activities outdoors. Children and the elderly should also avoid outdoor physical exertion.	**Reduce** or reschedule strenuous activities outdoors, especially if you experience symptoms such as coughing and throat irritation.

* People with heart or breathing problems are at greater risk. Follow your doctor's usual advice about exercising and managing your condition.

Figure 1.3: Air Quality Health Index (AQHI) and the Health message

The value of AQHI can belong to any one of the four health risk categories as shown in figure 1.3: AQHI value from 1 to 3 - low health risk, value between 4 to 6 - moderate health risk, AQHI value from 7 to 10 - high health risk, and value above 10 - very high health risk. There are also health messages associated to each category for alerting the public about the potential health effects when exposed to poor air quality shown in Figure 1.3.

The value of AQHI gives the risk level based on the exposure to pollution level as people get affected when they are exposed to pollution. In this way by identifying the value of AQHI the people can avoid the short-term exposure to pollution and help to build community with better air quality.

1.3 Air Quality measurement in British Columbia (BC)

Air quality monitoring in BC is measured by the provincial government, Metro Vancouver along with Environment Canada and regional districts. There are around 150 stations for air quality measurement and these are maintained by Ministry of Environment (MOE) and some industrial employees with permit. The main pollutants of interest in BC are Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Ozone (O_3), Particulate Matter ($PM_{2.5}$ and PM_{10}), Sulphur Dioxide (SO_2), and Hydrogen Sulphide (H_2S). These pollutants are measured in three ways: continuous monitoring, non-continuous monitoring, and mobile monitoring [38].

1. Continuous monitoring: This is an automated mode of measurement in which air quality is measured by drawing air through tubes. The collected air is automatically monitored, measured, documented, analyzed and validated. This data is then further checked for any errors and made available to public in an hourly manner in the current air quality data page by MOE.

2. Non-continuous monitoring: The non-continuous monitoring is also called as manual sampling. In this method the field technicians assigned by the ministry collects the air sample by placing filters or canisters for a discrete period of time (such as one, three, or six days). The sampling follows strict rules set out by the Ministry of Environment

(MOE). The collected sample is then sent to the certified labs where it is weighed and examined to understand the content and then the information is uploaded to the database of MOE. The manual instruments used to collect the sample values are Single Channel 16.7l/m (PM2.5 and PM10) monitors, Dichotomous (Coarse and Fine PM) monitors, Speciation monitors, Volatile Organic Compound monitors (VOCs), PAH monitors, and Passive samplers.

3. Mobile monitoring: The last method for collecting the data is by installing monitoring instruments in a large vehicle or an airplane. These instruments move around for short period collecting data around areas where fixed monitoring is not available. One such air quality monitor is Mobile Air Monitoring Laboratory (MAML) [4] build on Ford F550 chassis which can measure both continuous and non-continuous measurement.

Figure 1.4: The Mobile Air Monitoring Laboratory Vehicle [4]

The common air pollutants measure by MAML are black carbon, Sulphur Dioxide, Nitrogen Dioxide, Carbon Monoxide, Ozone, and Particulate Matter (PM). It also measures meteorological data like wind speed, wind direction, temperature, and humidity. The collected data is then analyzed and transferred to the database of the ministry.

1.3.1 Pollution Monitoring System in Prince George

The city of Prince George is located in the central part of the British Columbia province at the junction between the Nechako and Fraser rivers. The city has reported poor air quality due to forest fires, road dust, geographical location, transportation, industries, and other factors. The air quality of an area can be directly related to human activities within the physical environment [15]. For this reason, it can be seen that there is a huge variation in the distribution of pollutants from region to region. For example; the presence of an industrial source such as a pulp mill in and around this area can highly affect the quality of air. Other factors include topography, atmospheric condition, and magnitude of emission from sources [39].

Figure 1.5: Map showing the location of the eight monitoring station in Prince George [5]

Having considered all these factors over the years in Prince George, Government agencies along with industrial partners funded the installation of eight fixed monitoring stations to measure the dominant pollutants in the area [40]. Figure 1.5 shows the layout of the eight monitoring stations located in the region as reported in the air quality report of 2016 [5]. The data collected from these stations are monitored by the Prince George Air Improvement Roundtable (PGAIR) which includes the Ministry of Environment (MOE), the Ambient Air Quality Monitoring Working Group, and other volunteers.

Of the eight monitoring stations installed, there is a core station that measures all six pollutants which includes Particulate Matter ($PM_{2.5}$ and PM_{10}), Total Reduced Sulphur (TRS), Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Ozone (O_3). The other stations only measure certain pollutants or meteorological data [5]. The core station which is called the Plaza 400 monitoring station is located downtown and here the air quality can be recognized as a combination of industrial, commercial, and residential emissions [40]. This station contains several instruments such as API model 400 ozone monitor for ozone, API NO_x monitor for Nitrogen Dioxide, SHARP model 5030 for $Pm_{2.5}$), TEOM 1400a analyzers for $Pm_{2.5}$ and $Pm_{2.5}$, and TRS samplers for Total Reduced Sulphur [5] [40]. These data are available in the Ministry Of Environment (MOE) [41] website and the data can be accessed and downloaded by the public. The collected pollutant values are used to calculate the AQHI value which gives information about the local air quality that can help in determining health effects.

1.4 Motivation

One of the important components in dealing with the issue of air pollution is to increase awareness among the public about the current situation and its impact so that they can act on it. The conventional method of monitoring air quality with the help of a few expensive and stationary monitoring systems typically installed by government agencies may not be effective. To achieve the goal of public engagement, pollution monitoring must become part of daily activity for everyone. For that, the devices to monitor pollution must be small, portable, inexpensive, and part of a regional, national, and global system. With the technological advancement of low-cost computing, communication, and sensing devices, and the revolution of open source software [42], we believe it is possible to build a pervasive air pollution monitoring system with the available sensors from the market and open-source software. Now the question is how to design such pollution monitoring systems faster and make them accessible to as many people as possible.

Achieving the above-stated goal requires a suitable system framework that can help to accelerate the process of the design and implementation of an air pollution monitoring system using the off-the-shelf hardware and building open source tools for representing the data collected from the sensors. Some recent attempts to build low-cost air pollution monitoring systems have been done, however, none of them are simple and easily replicable. This thesis is an attempt to fill that gap by first proposing a simple and comprehensive framework and then demonstrating its feasibility and use by creating our own low cost and easy to use a pollution monitoring system that is operational in our lab. We have also added a step of calibration by implementing a web-based tool to ensure that we measure high-quality data. Our contribution is a step towards inspiring and motivating not only the public to use the device but also enable many amateur electronic hobbyists to buy and construct the hardware and download the associated software to build their pollution monitoring device.

1.5 Thesis Contribution

There are three major contributions from the thesis:

1. Air pollution monitoring system: The system itself which measures the pollutants from the atmosphere. This includes the sensors with the processor and the data transferring module. We believe that this could be a way to show that a low-cost system could be used for data collection.

2. Air pollution visualization software: The next major contribution is the software that could be used for data visualization. The main idea is to make the collected data user-accessible and hence we came up with the idea of building the complete tool from the scratch as the other available tools in the market are costly.

3. Calibration tool: Development of a web-based tool for sensor calibration to ensure the quality of data obtained from the sensor system.

1.6 Structure of the Thesis

The rest of the thesis is organized as follows. Chapter 2 focuses on a review of related work for different methods of understanding and detecting pollution. This is divided into four main categories and work done in each category is explained further. Next, in Chapter 3 the design and implementation of a pollution monitoring system and the visualization tool are explained. To test the accuracy of collected data to the original data we have implemented a calibration tool by linear regression which is described in Chapter 4. In Chapter 5, we present an analysis of the results obtained from the system. Finally, in Chapter 6 the conclusions and directions for future work are discussed.

Chapter 2

Literature Survey

The development of Wireless Sensor Network (WSN) is considered to be one of the greatest innovation in the field of electronics. The miniaturization of the components has allowed exploration of applications in various fields such as health care, military applications, traffic control, monitoring, and data collection [43] [44]. Out of all the applications, urban air quality monitoring has gained a lot of attention as it is one of the major issues faced by society today. While reviewing the related work, we found that there are various methods that can be used to understand pollution. These could be classified into three based on literature as follows [45] [46].

1. Vehicle-based sensor network

2. Community sensor network

3. Static sensor network

Further, we will be highlighting the important research done in each of the above four categories and will be classifying to which category our work falls.

2.1 Vehicle based sensor network (VSN)

In recent times, the number of private vehicles on the road has increased in proportion to the increasing population around the globe [47]. Even though the increase in the number of automobiles is one of the major factors that is contributing to the increase in pollution, certain researchers took this as a medium for measuring air pollution data. In this category of work, the vehicles (like buses or cars) are installed with a portable, low-cost sensor to obtain spatially resolved data.

One of the best ways to study the air quality is by collecting fine-grained data also called 'micro-climate monitoring'. However, since the existing monitoring systems are bulky and expensive it is impossible to obtain spatially resolved data. To solve this issue, in 2009 a group of researchers used mobile monitoring as a method and proposed a vehicular wireless sensor network [6] that measures the changes in concentration of a single pollutant (Carbon Dioxide in this case) by mounting the sensor node onto a vehicle. The system is equipped with a Carbon Dioxide (CO_2) sensor, a Global System for Mobile (GSM) module, a GPS receiver, and a ZigBee module to create an intra-vehicular network. The collected data is transferred through GSM short messages to the server and is displayed on Google Maps for results. The architecture of the VSN is shown in the figure2.1.

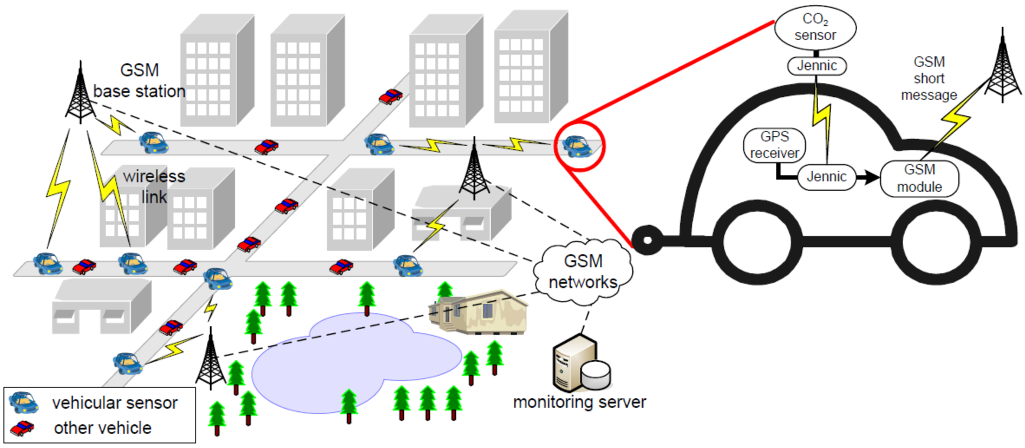


Figure 2.1: Architecture of Vehicular based sensor network [6]

This work faced two network-related drawbacks. Firstly there was a duplication of air pollution data as the number of vehicles in a given area changed dramatically from time to time and secondly, to reduce the data transfer in an area with many sensor nodes at the same time by exploiting opportunistic communication. These two problems were later addressed in 2011 by designing two message-efficient algorithms [48]. The first algorithm works by dividing the sensing field into a fixed number of grids and each grid was allocated with a particular reporting rate which is called dynamic reporting rate to reduce communication overhead. The second algorithm allows the sensing nodes to communicate with each other and find out their reporting rate and opportunistically transfers the collected data.

The next work discussed below involves data collection that requires more frequent and spatially dense pollutant measurement and is called a fine-grained approach. In the reference paper [49], the vehicular-based approach of measuring fine-grained air quality in real-time was demonstrated. To increase the spatial density, the study proposed two cost-effective data collection models: one for public transportation infrastructure and the other for a personal sensing device. In the public transportation infrastructure model, a Mobile Sensing Box (MSB) was installed in public transit buses which contained the microcontroller (Arduino), sensors for measuring Particulate Matter (PM), and Carbon Monoxide (CO), a GPS module, and a cellular modem. The module was powered through the bus batteries and the collected data were transferred to a server and visualized in a platform called Google Fusion tables interfaces. In the other data collection model, a Personal Sensing Device (PSD) that included an air quality sensor was installed in cars and connected via Bluetooth to a smartphone. This gave the user information about the air quality while driving through a specific area. The data obtained from both the models were compared using linear regression and the results showed a positive linear relationship between the data sets. These collected data were available to the public through customized web and mobile Apps.

A research group from Japan, Shirai et.al [50] proposed an effective method to acquire air quality data in urban areas in which the air pollution system has a sensing unit mounted on a public vehicle such as a garbage truck that moves in and around the city. The main focus is on pollutants like Particulate Matter (*PM*), Carbon Monoxide (*CO*), and Sulphur Dioxide (SO_2). The system is also equipped with a GPS module to identify the location. A control centre tool was developed to check for maintenance and remotely control the sensor system. The control centre consists of a map that tracks the route of the vehicles and the sensor data acquired by each vehicle. A monitor was developed along with the system to send the users (which in this case is the local citizens in Fujisawa City, Japan) the collected data. The system estimates the amount of pollution inhaled by the user by acquiring the user's location from the mobile application and mapping it to the location-sensors value which has been already computed.

Another mode of transportation that has been used to understand the air quality was the public bicycle system. The bicycle-borne sensor [51, 52] system consists of an exhaust gas sensor, Particulate Matter (*PM*) sensor, GPS module, Bluetooth module, and a microprocessor. The system collects pollutant data along with the location data from the GPS module and is stored in a storage chip. When the subscriber returns the bicycle to the dock station, the data is transferred to the data center via the Bluetooth module and is then visualized using the 'Baidu'heat map (similar to Google maps that are used in China).

Unmanned Aerial Vehicle (UAV) that do not need on board human pilot are also used to understand air quality [53]. This work focused on six different pollutants that contribute to Air Quality Index by mounting the sensor board on the UAV. The flights were carried out for 30 days with 20 minutes of monitoring. Along with the sensor board, a smartphone is attached that collects data from the air sensors by establishing a Bluetooth connection. The collected data is transferred to air quality analysis software that will display the real-time monitoring value along with AQI value.

The previous studies in this section show that collecting data using VSN can support micro-climate monitoring. The advantage of having spatially resolved data is that it helps to understand the local trends of pollution. Having high mobility of sensor nodes helps to cover a larger geographic area and gives access to data in the areas where the reference system is not available. However, the majority of work only focused on data collection and optimization of data. The main idea of deploying any system for measuring air pollution is to educate citizens and make them aware of air quality through the indices. The efforts to calculate AQI or AQHI was not in the studies described above. The communication of these indexes will make people aware of the air quality, but most of the work to date in this category has failed to bridge the data gap.

2.2 Community sensor network (CSN)

Technological development has paved way for a novel paradigm known as crowdsourcing. Crowdsourcing involves collecting information from a large group of people through the internet or other means and utilizing the collected information towards a common goal. In a community sensor network the use of crowdsourcing or participatory sensing is achieved with the help of a portable sensor network. This allows any citizen to collect data and transfer it to a common platform like a web interface. Various research work falls under this category.

One of the popular methods to achieve CSN is through mobile participatory sensing in which individuals carry portable handheld devices that measures the pollutants. Intel has developed a prototype system named 'Common Sense' [8] which is based on mobile participatory sensing that enables citizens to collect pollutant data. The system includes a handheld device that measures several pollutants (example: Carbon Monoxide (CO), Nitrogen Oxide (NO), Ozone (O_3), temperature, and humidity) and uploads the data using Bluetooth or GPRS radios for visualization over the web. This system was further tested by deploying it on a municipal fleet of street sweepers in the city of San Francisco [54].

The next portable and crowdsourced monitoring system on list is 'Mypart' [7] which is a wrist-worn particle sensor that measures particulate matter of 10 microns or less (PM_{10}). The design of MyPart is based on a laser-based photodiode system with integration of structural design and circuitry for effective visualization, BLE transceiver for low power networking, and also a mobile application for visualization.The two main issues tackled by MyPart are accuracy and calibration of the sensor, which no existing consumer sensor had addressed.

Another participatory-driven sensing system is ‘Eco-mini’ [9] which also a wearable stand-alone device for clinical use that measures Ozone (O_3), Sulphur Dioxide (SO_2), Volatile Organic Compounds (VOCs), sound level, temperature, and humidity values. This system is based on a low power microcontroller and consists of a GPS module for location and a Bluetooth module for data transfer. They developed webserver and a mobile application for data visualization. The next wearable work is ‘CitiSense’ [55] which is a system attached to a bag stripe which measures the air pollutants Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), and Ozone (O_3) along with environmental parameters such as temperature, humidity, and barometric pressure. The collected data from the sensor is processed by a microcontroller and transfers the data using a Bluetooth module to a smartphone which does the data storage, analysis, and data aggregation. The collected data is then transferred to a back-end webserver from where the user can get a personalized view of their data.

Portable Handheld Device   Wrist worn Particulate sensor   Wearable sensor monitor

Figure 2.2: Examples of Community Sensor Network [7], [8], [9]

There are also crowdsourced projects initiated in Vancouver such as TZOA [56] that can be clipped to the clothing and measures *PM* values and displays this in an application. These devices decrease the gap between individual and their awareness of polluted air in their local environment. In NewYork a project named ‘Aircasting’ [57] provides the health and environment data to the users with the help of the Android ‘Aircasting app’. The ‘Aircasting’ [58] platform includes a palm-sized monitor that measures $PM_{2.5}$, relative humidity, and temperature. Outside air is drawn through a sensing chamber and the particles are measured through the light scattering method. It also includes an LED wearable apparel named ‘Air casting Luminescence’ [59] that illuminates LEDs according to the real-time sensor measurement; varying from red for high intensity, then orange, then yellow, and finally green for low intensity.

The wide availability of the Micro-Electro-Mechanical System (MEMS) and Wireless Sensor Network (WSN) have changed how physical world collect data and interpret them. ‘G-Sense’ [60], for Global-Sense, is an initiative from the University of Florida in which they combine features of sensing platform applications like Location-Based Services (LBS) for tracking and location identification, Participatory Sensing (PS) for determining pollution index, and other environmental data, and Human-Centric Sensing (HCS) for health-related data for a specific group of users. The sensors collect data and send it to a first-level integrator where all the data is combined and then a data transport network transfers the data to the server that stores and performs data processing. The data visualization takes place from the server. Later another system which is considered to be the subset of ‘G-Sense’and was named as ’P-Sense’ [61] or Pollution-Sense. The architecture of this system is based on ‘G-Sense’in which external sensors are integrated using an Arduino development board. In this system the data collection is based on Participatory Sensing (PS) and the goal is to provide government officials, doctors, and community developers with data so as to get a deeper understanding. They have also pointed out the research and implementation challenges that need to be addressed when building a community networked system in considering issue such as security, privacy, and data visualization.

The work discussed in this category seems to be promising but at the same time the quality of data obtained, getting public involvement for data collection, and privacy issues [45] are a few of the challenges researchers are trying to deal with. The cost of maintenance of such a community network is also a crucial factor.

2.3 Static sensor network

In this category, the system is kept at a fixed location (example: traffic lights, street lights) or any planned areas [46] which collects the pollutant values and transfers the data to a visualization platform where the users can view it. These systems are designed to be inexpensive and so can be easily replicated or replaced. The system can be used for measuring either indoor or outdoor pollutants. There is a variety of research work done under this category and here I reviewed few relevant ones.

A research group from Hong Kong developed an Integrated Environmental Monitoring System (IEMS) [62] that integrates different environmental detection sensors ($PM_{2.5}$, UV sensor, noise sensor, temperature and humidity sensor) into a single system, and data from this system is used for processing and visualization. IEMS consists of Integrated Environmental Monitoring Devices (IEMD) which combine microcontroller units, sensors, and wireless communication modules. The research group also developed a Handheld Remote Control Panel (RCP) for the system which is an Android application that acts as an interface for the device control and handles the data exchange between IEMD and web server. Finally, the webserver provides real-time data visualization and data analysis. These systems were placed at bus stops, bridges, and even at construction sites.

Another research team from Mauritius developed a Wireless sensor network Air Pollution Monitoring System (WAPMS) [63] that utilized a data regression algorithm called Recursive Converging Quartiles (RCQ) to remove duplicate data and then calculate AQI values. The array of sensor nodes collects the pollutant data and transfers it to cluster heads where the RCQ is applied to improve efficiency and alleviate the congestion problem. From the cluster head, the data is sent to the server and represented (using line graphs for each area).

'AirSense' [10] is an approach to assess indoor air quality. This research work introduces the idea of indoor air quality by proposing a system which measures indoor pollutants. The system works through electronic sensors that are coupled to an Arduino microcontroller. The system not only extracts the data but also provides its users with effective visualization and analysis of the data. The researchers have developed this system to sense pollution and provide education and awareness among its users. This system made use of machine learning algorithms to predict the pollution sources and forecast their behaviour that in turn increases the predictive intelligence of the system. The system also uses a smartphone application that gives users an interface for visualization and interpretation of the data.

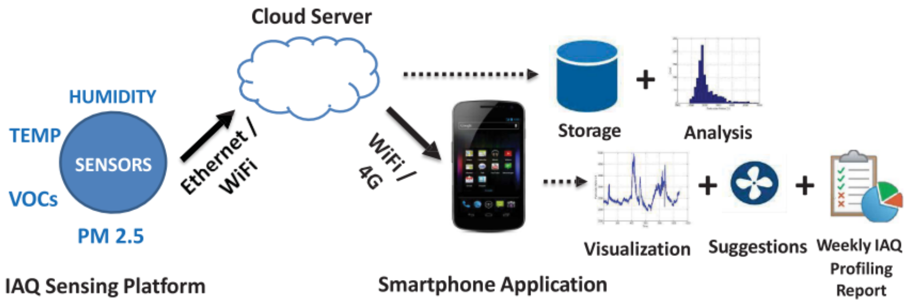


Figure 2.3: Architecture of a Static Sensor Network (SSN) - AirSense [10]

A different group of researchers from China also developed the system 'Air-Sense' [44] to monitor and predict the quality of air using the ZigBee network to connect the sensors. The system uses four different types of sensors: humidity, temperature, $PM_{2.5}$, and Total Volatile Organic Compound (TVOC includes the general organic gases). A ZigBee trans-receiver is used for communication with network nodes. This prototype is tested in different areas in the house.

Another static system that focuses on indoor air quality in which the main focus is to understand the pollution in an office environment where the pollution is triggered by electronic devices and machines [64]. In this, the pollutant measured is ozone, which is mainly emitted from a photocopier machines. The system is designed with different nodes where the sensing node contains the Arduino microcontroller which collects the data from the ozone sensor. The measured data is transferred through a Bluetooth link to a gateway node from where the data is forwarded over the Ethernet network to the processing node. The data is saved in a database and using a 2D graph the concentration gets visualized. A research group from Harvard University developed a wireless networking testbed called as 'CitySense' [65] in which multiple environmental sensors are attached to street lights. These sensors were deployed in Cambridge MA and data was uploaded to a server using mesh networks like RoofNet [66], TFA [67], and CUWin [68]. Using a web-based interface the data can be pulled from the server and made available to end-users. The main feature of this system is that the sensor nodes are powered from street lights and there is no constraint from battery life.

Liu et al [69] developed a micro-scaled air quality monitoring system for understanding the CO emissions from vehicles by integrating the sensor nodes with a network gateway. The data collected from the sensor is transferred to the gateway using a ZigBee communication link and from here meteorological data and collected sensor data are forwarded to a central system through GSM. This centralized control system is supervised by a LabVIEW [70] program which helps in store the data into a MySQL database. They deployed the system on the main roads of Taipei (Taiwan) and obtained accurate values of pollutant concentration.

2.4 Summary

In this chapter, we discussed the relevant works done in measuring the air pollutants using sensor networks. The work can be categorized mainly into three on the basis of carriers of the sensor node which are vehicle-based, community-based, and static sensor network. There has been a lot of research work done in each category and efforts are made to understand the air pollution by building a system and deploying it on a buildings, vehicle or by crowdsourcing. Our research project falls under a static sensor network category in which we have tried to integrate a system that measures the major pollutants in the city of Prince George and also provide AQHI values. Unlike the other systems mentioned in the literature, our main focus is to give user-specific data by categorizing the users into three; layman, data scientist and the public officials. We have also tried to implement a calibration procedure to ensure the quality of data. In next chapter, we will discuss the system design and its working.

Chapter 3

Design and Implementation of Air pollution Monitoring System

The use of complex, expensive, and stationary monitors for collecting and analyzing pollutant data began soon after the 1970 Clean Air Act came into action in the USA [71]. Ever since then researchers and environmentalists have been interested in studying the pollution data and its effects. This became even more popular after the development of low cost, easy to use, portable sensors in the market [72]. In our research work, we have attempted to fill the gap for understanding the quality of the data obtained from low-cost sensors by comparing these with the official monitoring system in the city of Prince George.

In this chapter, we share our hands-on experience in the design, development, integration, and operation of the air pollution monitoring system using low-cost sensors from the market. We have also build a visualization tool for effective data representation which is explained at the end of the chapter.

3.1 Design Goals

For the development of a reliable and simple pollution monitoring system, these are the steps we followed to make the system efficient. These goals are very unique to our work and are listed as follows:

1. Sensor Selection

   There are different kinds of sensors available in the market for measuring each pollutant. Selecting the right kind for obtaining accurate data was one of the challenging tasks. We referred to the Air Sensor Guidebook by US Environmental Protection Agency (EPA) [73] that came with guidelines that we followed before purchasing the sensor. One of the main factors that we looked at in the sensor is for accuracy. Keeping this in mind, we researched different available sensors in the market and selected the ones which we felt would meet our needs.

2. Processor Platform Selection

   To make the system work there needs to be a processor and platform. Choosing which processor to work with was the next task which we dealt with by answering the questions like:

   - Whether we wanted our system to be simple or complex?
   - Whether it is open-source in hardware and software?
   - The ease of programming
   - How easily can the processor be available?
   - What is the cost?.

   By answering each of these gave us clear criteria for selecting the platform.

3. Communication Module

   Once the sensors and platform were identified the next step was to find out a way how the data collected from the processor should be transferred to a database. We wanted the system to be wireless and for that, either Bluetooth or Wi-Fi module can be used. We wanted instantaneous data representation and the Wi-Fi module was a better choice for that purpose.

4. Easy Integration

   The integration of sensors with the Arduino platform is the next important factor that needs to be addressed. Some sensors can be easily integrated with any processor but others need driver code to be written to work with the processor. Also, the presence of one sensor should not affect the other sensors, which will result in data integrity problems.

5. Printed Circuit Board

   The initial circuit was built on a wired breadboard to create a prototype of the system. By working on a breadboard made it easy to experiment with different sensors. Once we have identified how each sensor should be placed in a breadboard arrangement the final version should be transformed into a Printed Circuit Board (PCB). By transforming the circuit design into a PCB makes the circuit stable, reliable, and robust.

6. Maintenance

   The next goal which we wanted to achieve was easy replication of hardware modules in case of any damage. For this we wanted the system to be "plug-and-play" model so that in case of any hardware issue, the debugging will not be challenging. We have also selected the hardware modules based on their availability in the market so that replacement will be quick.

7. Low Cost

   The final factor is cost. The price range for the hardware modules in the market varies from $20 to $800 per module in the category of low-cost sensor networks. We wanted to create a low-cost system that is affordable and set a target of under $400 for the entire pollution monitoring system.

3.2 System Architecture

In this section, we describe the design process for a low-cost air pollution monitoring system which can be categorized into system hardware architecture and system software architecture. The system is designed in such a way that the processor module collects the data through the sensors and transfers it into a database from where it is visualized. There is also a calibration procedure required which will be explained in the next chapter.

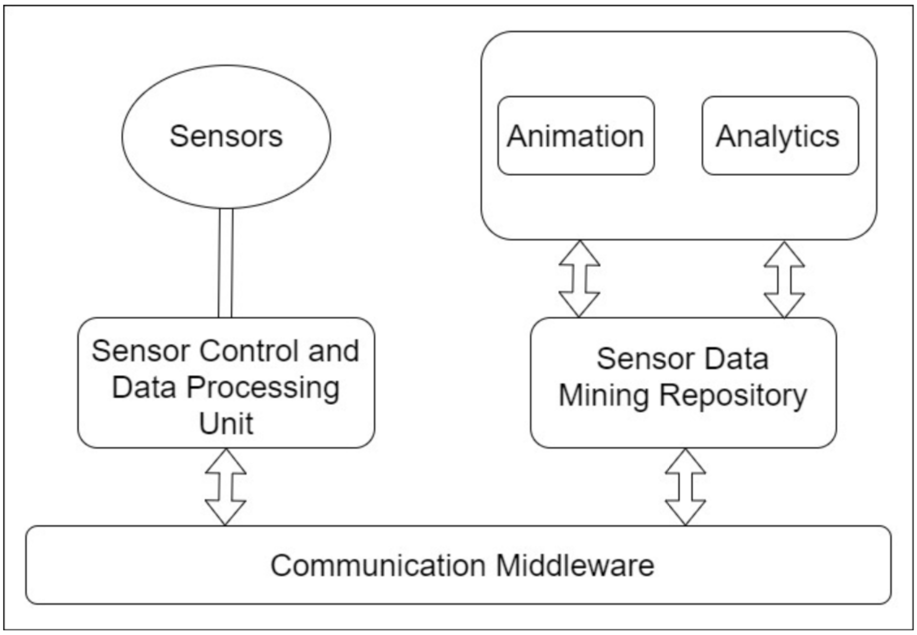


Figure 3.1: System overview

The system hardware section includes multiple sensors, the processor, and also the wireless communication module for transmitting and receiving signals. We will also discuss the air pollution visualization tool which mainly focuses on the representation of the data. The overview of the system is as shown in figure 3.1 and each part along with the sensor specification, implementation, and design will be discussed further in the section.

In this research project, we are measuring the pollutants in the city of Prince George.

The main pollutants of interests are $PM_{2.5}$, PM_{10}, O_3, NO_2, and CO. We have also measured the environmental variables like temperature and humidity along with the pollutants.

3.3 Hardware Architecture

3.3.1 Sensor Control and Data Processing Unit

This is the main hardware component for pollution monitoring as it is where all the other sub-modules are connected including the communication middleware. The main function handled by this module are as follows:

1. Control the sensors in collecting data.
2. Filter and process the collected data
3. Communicate and forward the data to the server.
4. Provide the necessary power for all the hardware connected to it.

For simplicity and ease of programming, we have selected a popular processor platform, the Arduino Ethernet board which has an ATmega328 microcontroller as shown in Fig.3.2.

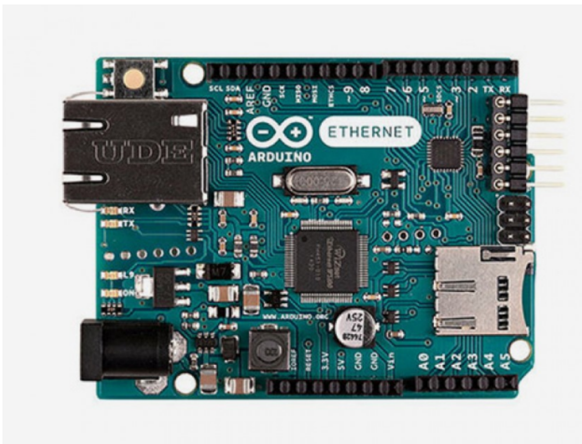


Figure 3.2: Arduino Ethernet Board

Arduino is an open-source physical computing platform that is divided into two parts, one is the hardware which is the board itself to which the external components are added and the other is the software which is the development environment for the programming. It is very simple to use the board with many external devices such as sensors or actuators and is widely used by researchers. There are different features which makes Arduino popular and can be listed as [74]:

1. Arduino is multi-platform and can be used with Windows, MacOS, and Linux.
2. The program is uploaded to the module via USB cable.
3. It has many I/O ports for interfacing external sensor hardware.
4. It is open-source hardware and software and the code can be easily downloaded.
5. There is an Integrated Development Environment (IDE) which is used as an interface for communicating with the hardware and is very simple.
6. There is an active Arduino forum in which many researchers and developers contribute their ideas and will help in trouble shooting.
7. The cost of the hardware is very low (under 60 CAD) and is easily available.

Description	specification
Microcontroller	ATmega328P
Operating Voltage	5V
Input Voltage Plug (recommended)	7-12V
Input Voltage Plug (limits)	6-20V
Input Voltage PoE (limits)	36-57V
Digital I/O Pins	14 (of which 4 provide PWM output)
Analog Input Pins	6
Flash Memory	32 KB (ATmega328P) of which 0.5 KB used by bootloader
SRAM	2 KB (ATmega328P)
EEPROM	1 KB (ATmega328P)
Clock Speed	16 MHz

Table 3.1: Technical specification

The board can be powered either by using a FTDI cable/USB serial connector, external power supply or using a optical Power Over ethernet module (PoE). The detailed specification [75] of the board is given in the table 3.1. These specifications give the freedom for researchers to explore different electronic devices easily. The board is programmed using the Arduino Integrated Development Environment (IDE) which uses an embedded C language.

3.3.2 Sensor

In chapter two we have described the various categories of sensor networks and how effectively these can be used in different platforms. In this section, we will be giving an idea of different sensors available in the market and the sensors which we selected for this research project. There are a variety of options available for sensors to monitor the various pollutants. One such category is the Metal Oxide Semiconductor (MOS) gas sensor also known as the semiconductor gas sensor, which is used to detect the concentration of hazardous gases in the atmosphere. The most popular series available in the market for this category are MQ-XX sensors which are popular for their wide detecting scope, long life, stability, high sensitivity, fast response, and also simple drive circuit [76]. The sensing material is made of either from Aluminium Oxide (Al_2O_3) or Tungsten Trioxide (WO_3) based ceramic and has a coating of Tin Oxide (SnO_2) that acts as the sensing material for the desired gas.

The sensing element is heated by Platinum wires which are connected to leads made of Nickel-Chromium. The gas to be detected has a specific temperature at which it is ionized and the task of the sensor is to work at that temperature. Once the gas is ionized it will be absorbed by the sensing material which changes its resistance and in turn changes the voltage across the sensor and can be read by the microcontroller [77]. The voltage value along with a reference voltage is used to determine the resistance of the sensor. Once the resistance of the sensor is known then by using a calibrated sensitivity curve, the concentration can be found.

Another popular MOS sensor that we explored was MICS, which is MEMS-based whose mode of operation is similar to the sensor above as both of them are metal oxide. Here, oxidizing gas or the pollutant gas adds to the insulative oxygen species causing the resistance to increase [78]. We also considered optical sensors which are spectroscopic devices which use light scattering principles to find the concentration of pollutants. These sensors are known for its capability to detect particulate matter of different sizes and is one of the recent innovations in the field of air quality monitoring. High responsivity, reliability, and long life are the main highlights of this type of sensor.

Selected Sensors

After understanding the wide range of sensor options in the market the sensors that were selected are listed below:

1. MQ-2 Sensor: This is a semiconductor gas sensor which has an electrochemical sensor which detects multiple gases such as Carbon monoxide, LPG, methane, and combustible steam. The sensor is connected in series with a variable resistor to form a voltage divider circuit and the variable resistor is used to change the sensitivity. The range of detection of gases is from 100ppm to 10,000 ppm and has a high sensitivity and fast response time. The sensor is small and portable and provides simple integration with the Arduino platform.

2. MQ-131: Sensor used for ozone detection. The operation of the sensor is similar to MQ-2 sensor. It decreases the resistance when exposed to ozone and becomes more conductive when exposed to large concentration of the gas. This can be used to measure the concentration of ozone in air. The detecting concentration range is from 10ppb to 2 ppm of ozone and also has a fast response time and long life.

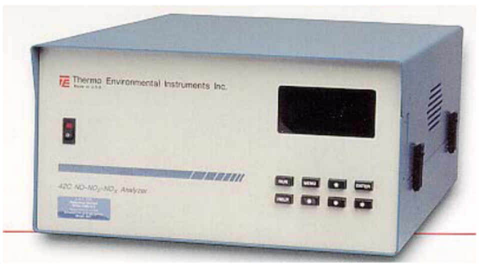
   The reference system in downtown uses an API model 400 ozone monitor for the city measurement [79]. The Figure 3.3 shows the picture of the Ozone sensor used for measurement at the reference station as well as in our low-cost sensor system.

API MODEL 400 OZONE MONITOR

MQ 131 OZONE SENSOR

Figure 3.3: Ozone sensors used for measurement

3. MICS-2714: This is a robust MEMS sensor used for the detection of Nitrogen dioxide (NO_2). The detection range is from 0.05 to 10 ppm and has a response time of 10 seconds. The sensor is comparatively small and of low cost. The reference system located in Prince George plaza 400 is API NO_x monitor [79] which uses chemiluminesence principle to measure the pollutant . The figure 3.4 shows the image of both the sensors for detecting the pollutant.



API NOx MONITOR

MICS 2714 SENSOR

Figure 3.4: Nitrogen Sensor used for measurement

4. PPD42NS Particulate sensor: The sensor detects particulate matter through the light scattering mechanism and consists of an infrared LED positioned at a forward angle to a photodiode. Variation in light density occurs due to particulate matter in the beam, and the photodiode detects this and changes the current from the diode [80]. The circuit generates a measurable signal which is proportional to PM concentration

[81]. This sensor can measure both PM2.5 and PM10 concentration alternatively. The reference system situated in downtown uses two different method for measurement, continuous and non-continuous [29]. Both the measurement method calculates the pollutant by drawing air from the atmosphere through an inlet and is placed on a Teflon coated glass fibre filter [79]. The Figure 3.5 shows the pictures of the sensors placed at downtown as well as sensor used for our system is shown.The comparison of both the measurement is done in the following section.

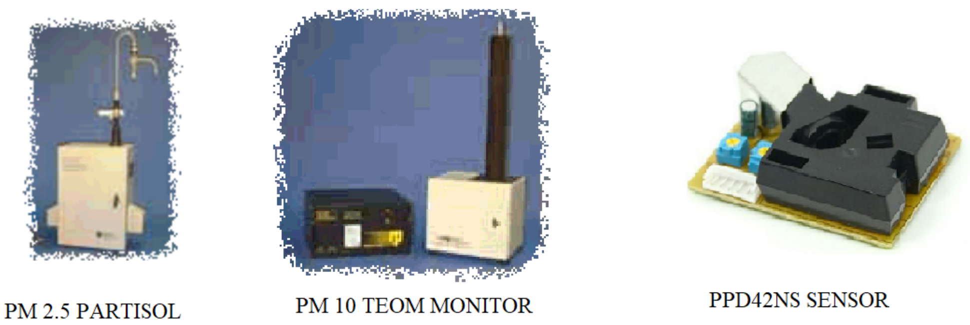


Figure 3.5: Particulate matter sensor used for measurement

5. DHT11: This is a very low-cost sensor available for temperature and humidity measurement and has a calibrated digital output. The measurement range for temperature is from 0 to 50 degrees Celsius. The device can be integrated with almost all Microcontroller platforms and is considered to be the best choice for many applications.

3.3.3 Communication Module

The collected data needs to be transferred to a server and from there to a visualization tool so that user can understand and interpret the data. Selecting the right communication module was hard as the main requirement was to establish a stable connection that will constantly send data over the network. At the initial stage of the research project we used an ESP module for data transfer. The WiFi module, ESP8266, is a highly integrated SOC that meets the requirement of efficient power usage, compact design and reliable performance [82]. The ESP module can be connected to a processor or it can also operated independently as the module itself acts as a microcontroller unit. Unlike the sensor modules connected to an Arduino module which needs a 5V power supply for its operation, this module needs a 3.3V power supply. The ESP module comes with installed firmware and it communicates using AT commands. On entering AT command in the serial monitor, the output would come as 'OK' if there is a successful connection.

However, this module failed to provide a reliable connection to the network and it also faced issues with power-up when all the sensors were connected. This became a challenge for data transfer so we chose a more efficent and reliable module called 'WEMOS D1 mini'. WEMOS is a miniature microcontroller based on an ESP8266 and can be easily integrated with Arduino. The development board has eleven digital input/output pins and one analog pins. The programming for the device is done through the Arduino IDE and flashed via the USB connector. The programming for the system was very similar to the earlier ESP module and it provided a stable connection.

3.3.4 System Overview

This section will show the integration of the above-mentioned sensors with the Arduino and communication module. After selecting the sensors and processor for the air pollution data measurement the next task was to put together a working system. This was done one step at a time so that troubleshooting will be simpler. The complete circuit diagram of how the components are integrated to the Microcontroller is shown in Figure 3.6.

Each of the sensors was carefully arranged to reduce spatial interference. The majority of the sensors use heating devices and heat discharges affects the operation of the sensors placed nearby. The initial circuit implementation was done using a wired breadboard. Building the circuit with a breadboard gave the flexibility to work with different arrangements.

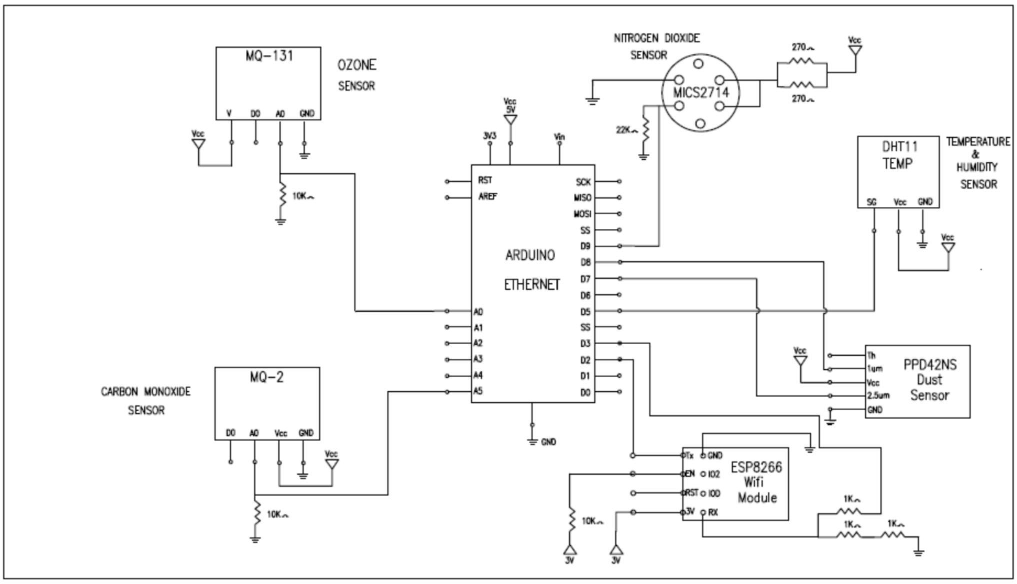


Figure 3.6: Circuit diagram of air pollution system

After building the circuit in a breadboard, it was migrated to a Printed Circuit Board (PCB) developed in Flemings Solution Ltd, India. The PCB was designed in such a way that each of the sensors can be plugged into the Arduino board which gave flexibility in case of any damage. This was done in case of any malfunctioning of sensors or processor, so could be replaced easily. The main advantage of using a PCB design is that all the components are fixed and there is no wiring at all. This produces a stable circuit for the sensor platform.

3.4 Software Architecture : Customizable Layered Visualization (CLV)

The hardware is responsible for the data collection but making the data available to users is done by effective visualization software. The main objective for developing the system was that the data collected should be accessible from anywhere. We developed a customizable layered visualization for the data that potentially involve different stakeholders. We believe that an effective visualization of relevant data is critically important for effectively combating major issues. The software design is driven by the following four key elements:

- Stakeholder specific visualization.
- Author-guided and user-driven interactivity.
- Hierarchical approach to visualization.
- Providing options for users.

3.4.1 Framework

The pollution monitoring stations collect huge amounts of data for pollutants and there are various ways to make people aware about this data. There can be different groups of users who will have access to the data which can be categorized as layman, data scientists (educators or researchers) who will be analyzing on the data, and the policy or lawmakers who will take necessary actions. Accurately identifying the stakeholders with an interest or stake in air pollution data is challenging. The paper on health, safety, and environment [83] categorizes stakeholders in environmental risk decisions into four categories of stakeholders as risk losers, risk gainers, risk perpetrators, and risk managers.

'Risk losers'are those who are susceptable to pollution which includes the general public or the society as a whole and 'risk gainers'are those who gain favorable outcomes from environmental risk decision mainly through economic gain. 'Risk prepetrators'are those who create risk and 'risk managers'are those who implement laws and regulations by looking at the trends of pollution for a certain period. In a similar way, we have also categorized our stakeholders into three main categories as layman, data scientist, and policymakers. We believe this categorization for representing air pollutant data is very efficient as it helps in representing what each stakeholder wants to view in the data.

The visualization should be simple for users and be able to drive the user to details of interest. It should not necessarily represent all the data in a single screen, which could be confusing to users and will lose the interest of different stakeholders. The user should be able to choose what data they need to view. If this approach is included in visualization design, the data that different stakeholders are looking for will be passed to them in an efficient manner.

3.4.2 Implementation of CLV of Air Pollution Data

The complete software implementation for the visualization software was created using a combination of Node.js and Python. The framework for the software is shown in Figure 3.7. In Figure 3.7 the back-end and the front-end of the software is clearly shown and how it is connected with each other. The front-end is the representation layer with which the user interacts to get an output, for example, the webpage which represents information demanded by the user. The back-end of software can be described as what makes the front end work and it deals with the server-side. The back-end mainly focuses on database, scripting, and architecture.

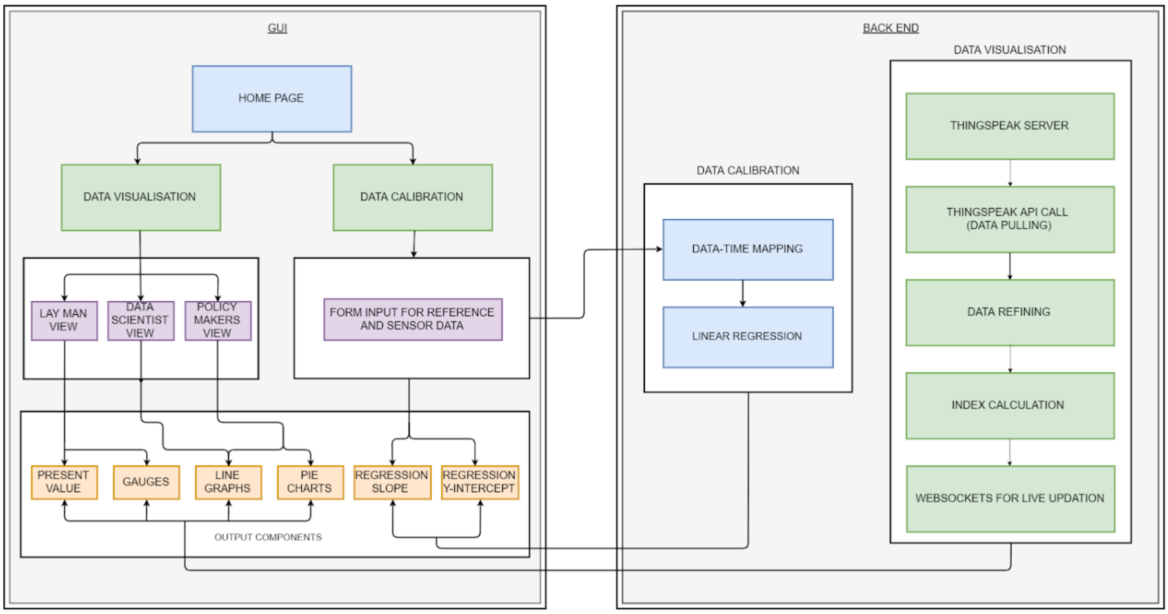


Figure 3.7: Software Architecture

According to the classification of stakeholders, each user group is looking for different views of the data. For example, the data could be represented in the form of huge spreadsheets or documents. There are different methods for visualization for large data sets, such as heat-maps, bubble plots, box plots, bar graphs or a plane graphs. The most common method for displaying complex data is a graphical method, which can show the data corresponding to a certain time or day. We believe that instead of visualizing the entire data set through graphs or plots, it would be easier for the layman to interpret the data if it is a single instantaneous value. The representation implemented for the lay man stakeholder is as in Figure 3.8, which shows a single value for the pollutant. The implemented representation of the data is easy to understand and it shows the concentration that each pollutant adds to atmosphere. The idea of this kind of visualization is that the user who falls into layman category simply needs to read the data from the display.

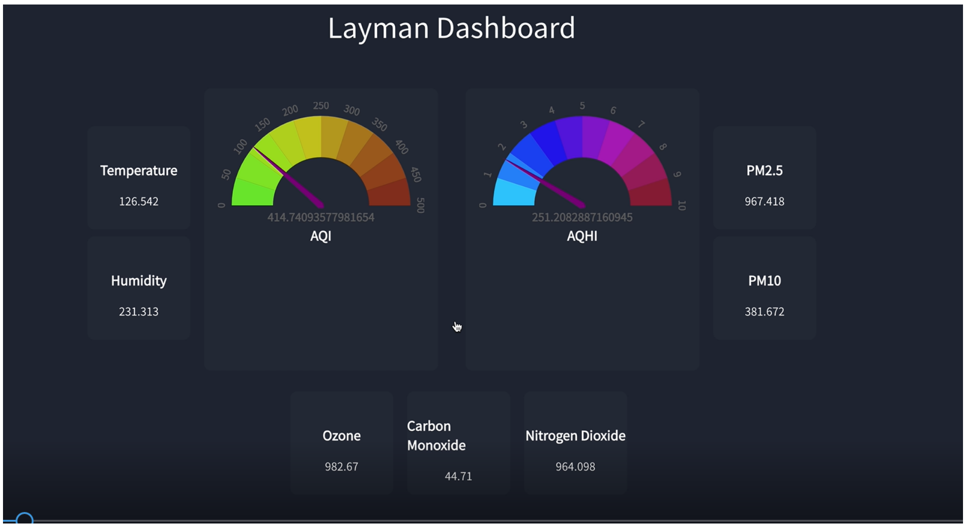


Figure 3.8: Dashboard for Layman category

The next data visualization implementation is done for the data scientist or researcher and is as shown in Figure 3.9. The ideal way to represent this more detailed view of information is through line graphs to show the values obtained over a range of time. The measured value of pollutants could be seen when the pointer hovers over the graph. These graphs are updated every time the source data is updated by the system, as researchers need to know real-time data to understand the scenario.

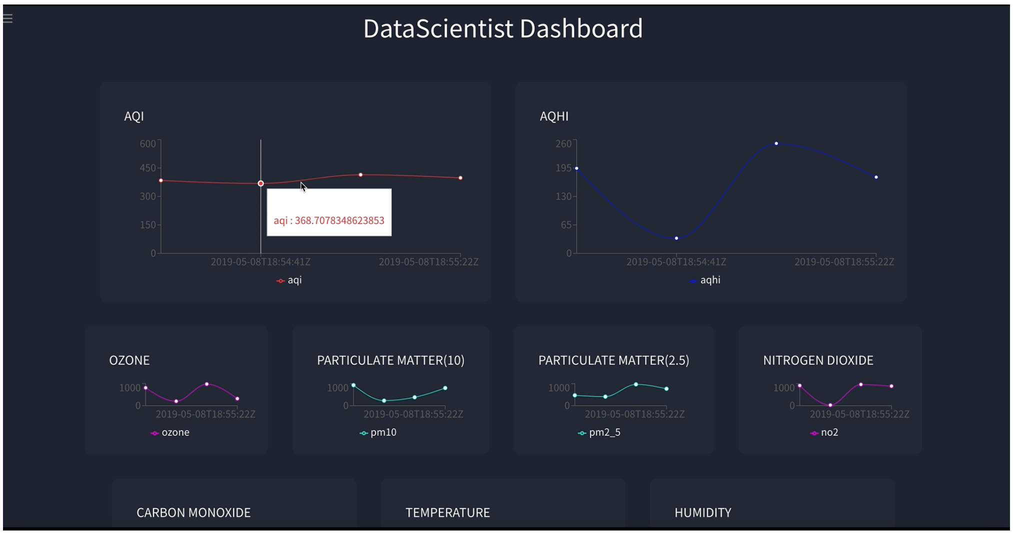


Figure 3.9: Dashboard for Data Scientist or researcher

The last categorization of users are the policymakers who implement actions based on changes in the concentration of pollutants. They are responsible for creating or enforcing policies and laws by looking at the pollutant data and hence they are the ones who need to see a combined view of all the pollutants and the trends in data as shown in Figure 3.10.

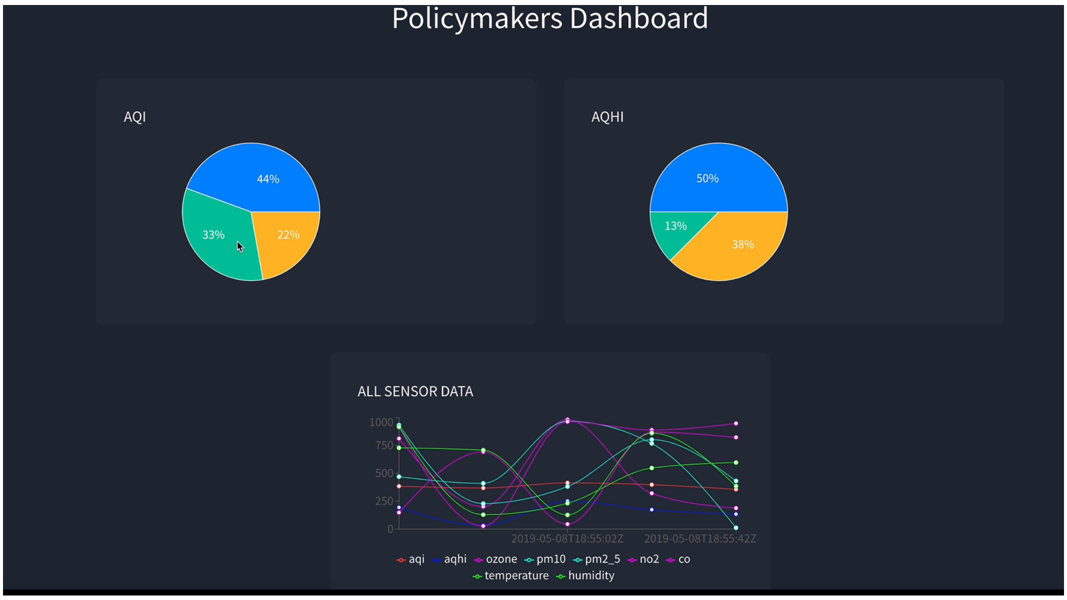


Figure 3.10: Dashboard for policymakers

This can be visualized by pie-charts which shows the percentage of each pollutant that contributes to the indexes. There is also a combined line graph that shows concentrations of different pollutants. By looking at the pie-charts and the combined graphs these users would be able to understand the trends and also could identify the pollutants that contribute more to the indexes.

The idea of such a hierarchial level of visualization is that it changes according to the users or groups who are focusing on these data. Albert Cairo, a renowned information designer claims that the fundamental goal of data visualization should be to make people better informed [84] [85], similarly we also believe that well communicated data is important for visualization.

3.4.3 Data Storage - ThingSpeak

The previous section demonstrated how the data can be visualized. One other aspect to be noted is where does the data get stored? The pollutant data that we collect from each system needs to be stored so it can be retrieved when needed. To implement this we use an open source IoT platform called ThingSpeak [86]. This is a visualization service that allows the system to aggregate, analyze, visualize, and store the sensor data [86]. ThingSpeak can allow many users to integrate their systems with other systems for collective and cooperative analysis of sensor data and promote citizen science in solving important local, regional or global problems. It offers a ‘hub’model for data repository and a set of APIs for accessing and using the sensor data for analysis and interpretation. We have used the data repository service of ThingSpeak (which is a free) to act as our server. The data transferred from WEMOS is stored using an API key provided by ThingSpeak and updated when new data is available. The data can be downloaded in the form of csv file which is simple to use in the subsequent analysis.

3.5 Summary

The implementation of a sensor system using hardware and software along with their integration have been explored. We have discussed the sensors selected and how these connect to the Arduino platform. The different stakeholders are identified and how the data is represented for each classification has been discussed. The next concern is the quality of data collected, which will be explained in the next chapter.

Chapter 4

Calibration Of A Sensor Network

The use of sensor networks for pollution monitoring has given the freedom to set up a monitoring system at residences, offices, or even at schools. One of the main drawbacks of this is to identify the accuracy of the data collected from these low-cost sensors compared to the reference monitoring system. If the sensor system gives values that are different from the reference system then it brings into question the advantages of this technology. This issue can be resolved through calibration which will reduce the uncertainty in data and make the output more accurate.

Calibration can be defined as an act of evaluating and adjusting the precision and accuracy of measurement equipment [87]. If the measured output from the sensor is not equal to the output from a reference system then it shows that there is a need for calibration. In general, all electronic instruments are calibrated according to given set conditions and acquire a certification of calibration before being sold. Even then the measured value might not reach the specified accuracy as the condition in the environment where it was calibrated changes, leaving the user with values that give poor information. This issue was taken up and explored by many researchers around the globe. During the 2013 Air Sensor Workshop, the US Environmental Protection Agency (EPA) suggested the three "straw-man calibration" approaches to improve the usability of uncertain data [88].

The first approach is ‘signal-based calibration’technique that requires the data from the reference station to be broadcast to the local station where the sensor is located and which will receive this data and perform a single point calibration of its response. This approach is feasible if the sensor is equipped with the data collection and can process automatic calibration.

The next approach for calibration is called ‘direct sensor calibration’that involves placing the sensor in a chamber in which a known concentration of pollutant is present and the response is observed. As the concentration of the pollutant is known, the output curve can be compared with it and calibrated accordingly. This is the most common method used for calibration in laboratories. Another way of approaching ‘direct sensor calibration’is by inspecting the pre-defined response given by the manufacturer and determining the sensor response to a given concentration. In either case the calibration requires equipment and skills to introduce an accurate concentration value.

The last approach is ‘secondary data normalization’in which the concentration values of the pollutant from the low-cost sensor are normalized in accordance with the federal reference method (FRM) or federal equivalent method (FEM) analyzers. This approach is cost-effective when compared to the other techniques and is less complex. This can be achieved by the use of a linear mathematical equation model that will convert the non-calibrated data into data of an acceptable form. The linear relationship equation $y = mx + c$ where m is the slope and c is the intercept of the sensor raw data is compared with analyzer data and a relationship pattern is established from this. The drawback of this approach is that the sensor does not always give a linear response and thus not applicable for all the curves. Even though these Straw-Man approaches were defined it can be hard to implement any of them in practice.

By the end of the workshop, it was widely agreed by the researchers that developing a software tool that can guide researchers through calibration procedures. By Fall 2016 the EPA developed a ‘Macro Analysis Tool’(MAT) that performs comparisons of low-cost sensor data with the reference system data and is included in ‘Air Sensor Toolbox’ [89]. In our research project, we replicated the MAT tool in the form of a webpage so that the user can upload data and recieve back the calibration curve. In the next section, we explain the working of the MAT tool and the procedures we followed for data calibration in detail.

4.1 Macro Analysis Tool - MAT

The development of the MAT supports researchers in calibrating sensor data. This is an Excel-based, user-friendly macro tool that compares sensor data with reference data [11] even if measurements weren't recorded at precisely the same time, or were collected at different time sampling intervals, such as 1-minute versus 5-minute intervals [90].

The MAT tool allows the user to insert the collected data and timestamps from both the reference system and low-cost sensor system into their respective pages of Excel sheets. After the data are provided the details of pollutants, time interval, measurement units and data completeness (amount of usable data obtained) are added in the control panel page of the MAT as shown in Figure 4.1. Once these are filled out the user clicks the 'RUN'button to perform the linear regression procedures.

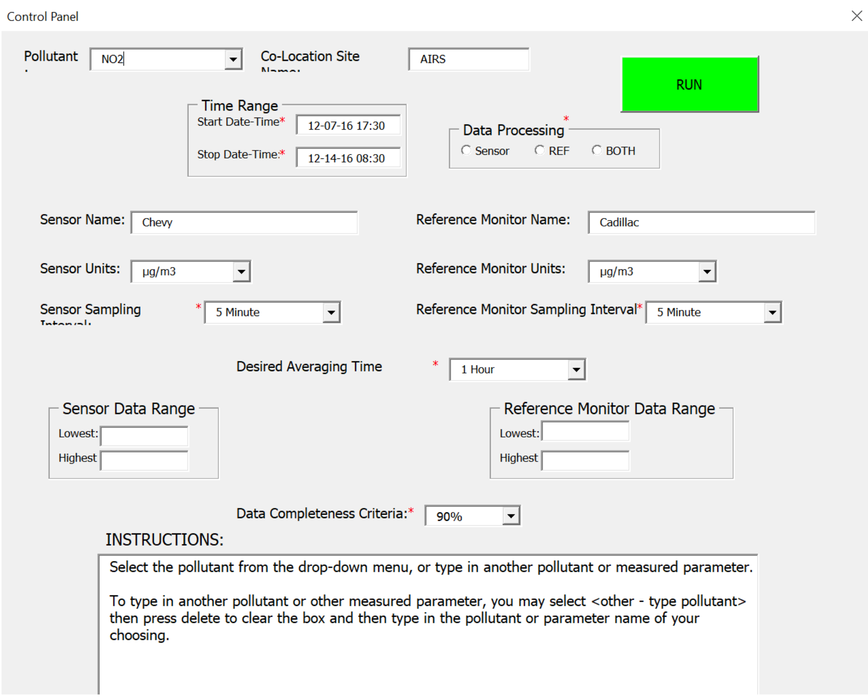
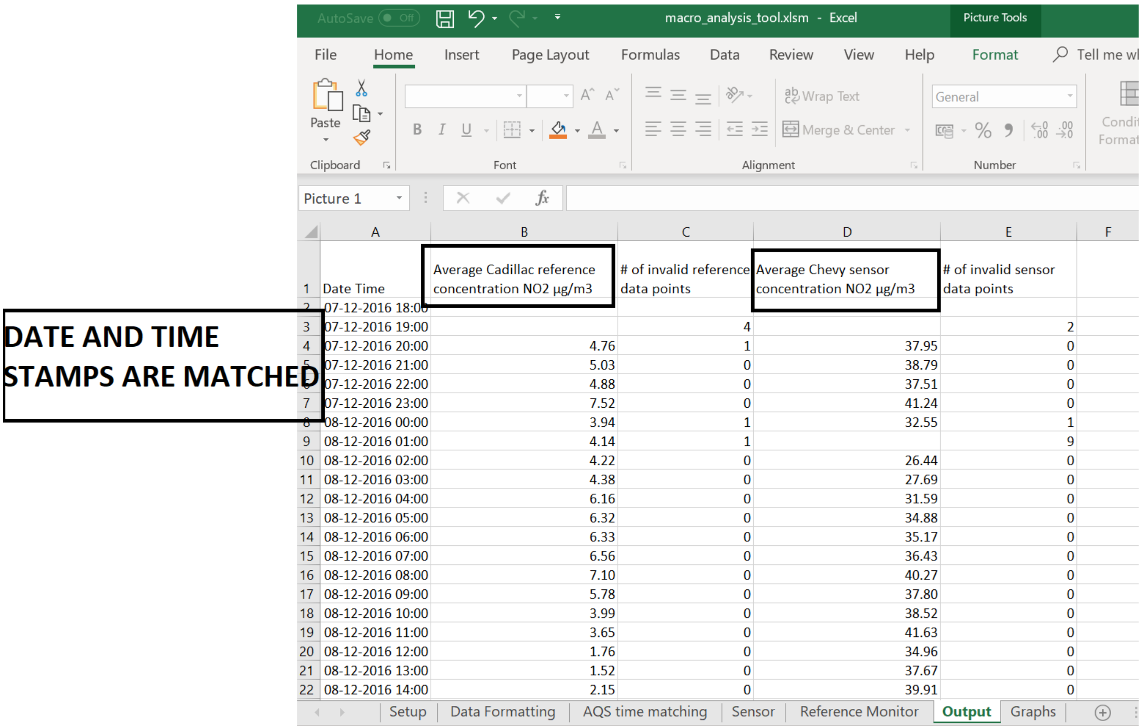


Figure 4.1: Control panel of MAT tool [11]

An example of the output page of the MAT is as shown in Figure 4.2 in which the two sets of data being compared as per the control panel setting in the tool [11]. This output page gives a statistical comparison of the two different data sets. From this page it can be seen that the date and time stamps are averaged to a single value for both systems. There is another column in the sheet which lists the invalid data points and gives the total number of non validated data points during the observation time. These non validated data points occur due to either data issues from the instrument or issues in reading values by MAT or unacceptable ranges specified by the user.



	A	B	C	D	E
1	Date Time	Average Cadillac reference concentration NO2 µg/m3	# of invalid reference data points	Average Chevy sensor concentration NO2 µg/m3	# of invalid sensor data points
2	07-12-2016 18:00				
3	07-12-2016 19:00		4		2
4	07-12-2016 20:00	4.76	1	37.95	0
5	07-12-2016 21:00	5.03	0	38.79	0
6	07-12-2016 22:00	4.88	0	37.51	0
7	07-12-2016 23:00	7.52	0	41.24	0
8	08-12-2016 00:00	3.94	1	32.55	1
9	08-12-2016 01:00	4.14	1		9
10	08-12-2016 02:00	4.22	0	26.44	0
11	08-12-2016 03:00	4.38	0	27.69	0
12	08-12-2016 04:00	6.16	0	31.59	0
13	08-12-2016 05:00	6.32	0	34.88	0
14	08-12-2016 06:00	6.33	0	35.17	0
15	08-12-2016 07:00	6.56	0	36.43	0
16	08-12-2016 08:00	7.10	0	40.27	0
17	08-12-2016 09:00	5.78	0	37.80	0
18	08-12-2016 10:00	3.99	0	38.52	0
19	08-12-2016 11:00	3.65	0	41.63	0
20	08-12-2016 12:00	1.76	0	34.96	0
21	08-12-2016 13:00	1.52	0	37.67	0
22	08-12-2016 14:00	2.15	0	39.91	0

Figure 4.2: Output page of Macro Analysis Tool [12]

The next output page provided by the tool is the linear regression result which is a correlation graph between the reference monitor and the sensor data. An example of the output page is shown in Figure 4.3. The graph drawn is a scatter plot and a slope-intercept line will be drawn through the data points. This line describes the average behaviour of the sensor data on the vertical axis compared with the reference data on the horizontal axis [11]. The fit of the graph shows the similarity in sensor measurements when compared with the reference instrument, on average.

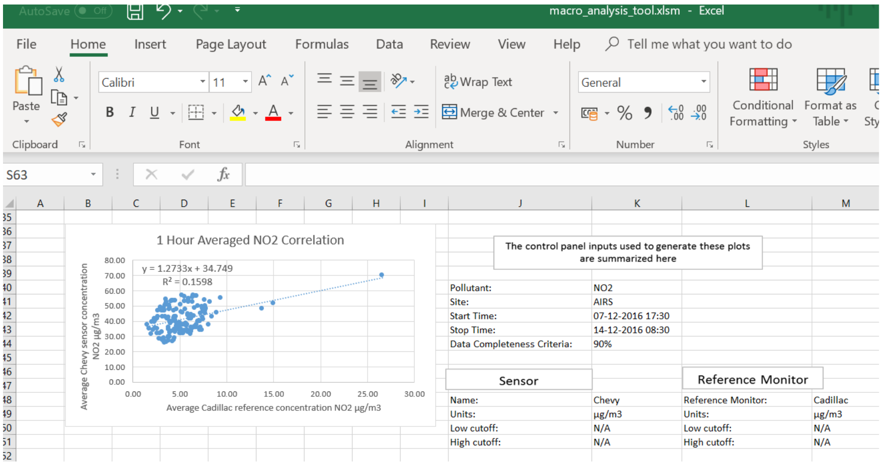


Figure 4.3: Correlation Graph

The tool also generates a correlation coefficient (R-Squared, R^2) that indicated how close the values are to the slope-intercept line. The R^2 ranges from 0 to 1 and the closer R^2 is to 1, the stronger the agreement between the sensor and the reference data [91]. Along with these outputs, the tool also generates a time series output graph which shows the concentration of pollutants for both systems as shown in Figure 4.4

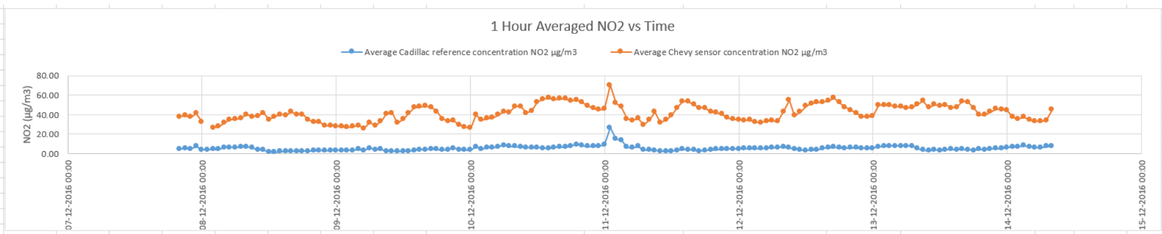


Figure 4.4: Time-Series Graph

From the time-series graph in Figure 4.4 it can be seen how the closely value from the two systems are related. Having included all of this functionality in the MAT gives a workable solution for the calibration issue. The development of the tool was a year-long project for EPA with community groups (Clean Air Carolina) and one tribal nation (Eastern Band of Cherokee Indians).

4.2 Calibration Procedure

The calibration procedure we followed is from the ‘Air Sensor Toolbox’guidelines for researchers, citizens, scientists, and developers who are working with low-cost sensors and their calibration to get more insight on air quality [89]. The Air Sensor Toolbox [73] provides a three-step procedure for calibrating a sensor:

1. Comparing the data from the low-cost sensor with a reference instrument.

   The data collected from the sensor system should be compared with the data from an already calibrated system that is placed by the local authority. This type of comparison is called ‘collocation’and to collocate, we need to find out where the reference system is placed and get access to the data from the reference system. Once we have access to the data collected from both the reference system and the low-cost sensor system, then the data can be downloaded in the desired format.

2. Creating a calibration curve with the help of the Macro Analysis Tool (MAT).

   The relationship between the response of an analytical instrument to the concentration or amount of an analyte introduced into a known instrument is referred to as the “calibration curve” [92] and is obtained by linear regression. Linear regression establishes a relationship between two variables, the independent variable which will be on the x axis and the dependent variable in the y axis by drawing a best fit straight line [13]. The MAT tool by EPA is used for finding the calibration curve and will give an error function as output. This error function is an equation which can be used for calibrating the sensor.

3. Repeating the calibration periodically.

   This procedure of calibration should be done periodically as the performance of instruments changes.It is expected that there will be changes in the best fit equation over time and this should be noted so as to get accurate values.

4.3 Framework

We believe that the MAT tool provides an efficient method for a calibration procedure for a low-cost sensor system. In our work, we have replicated the MAT tool into an IoT platform. This platform gives users the ability to input the values collected from the reference system and any low-cost sensor system. Our implementation makes the calibration procedure possible for any internet connected sensor system, so does not require the Excel software product.

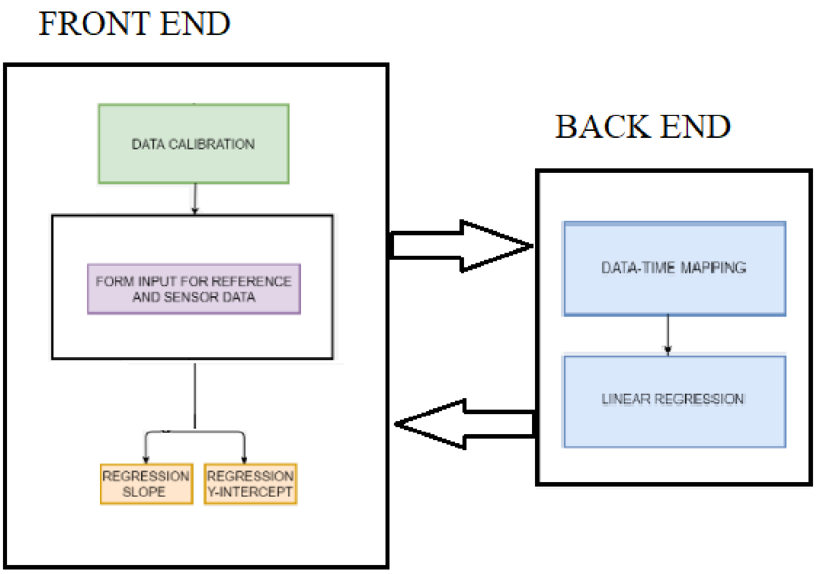


Figure 4.5: Framework of Calibration tool

Data is uploaded to the online platform in a comma-separated values (csv) file format along with their date and time. After the required data is uploaded the linear analysis takes place and the regression slope and y- intercept is provided as the output. This output regression equation is then used as the calibration equation and applied to the system. The framework of the calibration tool is shown in Figure 4.5. This tool can be accessed from anywhere in the world. The main reason for building such a tool was to have it easily accessed by any networked IoT device.

4.4 Summary

In this chapter, we have discussed how we can establish with the accuracy of our low- cost sensor system. We have discussed the MAT used for calibrating the measured data from the system and we have implemented an IoT version of the tool. In the next chapter, we will discuss the results obtained from our system and its comparison with the reference system in Prince George.

Chapter 5

Experimental Evaluations

In chapters 3 and 4, we have discussed the architecture and the calibration applied to the system. The main objective of the study was to compare how closely the calibrated data values from the low-cost sensor network are to that of the reference system located at the Plaza 400 building, downtown Prince George. The sensor system was deployed for six days from 30 May, 2019 to 4 June, 2019 and the calibrated data was collected and later compared to the values from the reference system. Before getting into the deployment section, we will briefly describe the complete system and will show how the data values are collected from our sensor system.

The Arduino microcontroller is connected to five sensors and these sensors collects the values for a fixed interval of time, in our case we have collected values for every five minutes and taken the hourly average of that data which gives 24 data points for a day. These sensor readings are digitized by the Arduino and a set of equations are applied to the raw signal and converted to concentration values. In the initial cycle, which is the calibration process, we collect these concentration values and supply them to the MAT tool for calibration. Once the regression equations are determined, we supply these to the Arduino to get subsequent calibrated values. The complete flow diagram of the system is as shown in Figure 5.1. Further, we will be discussing the comparison between the calibrated values from the sensor system to the data collected from the reference station. These are the observations that we have obtained after we have put our system out for six days. We have also added the weather data which includes the temperature, humidity and sunlight data in Prince George during the measured days in the Appendix section for reference.

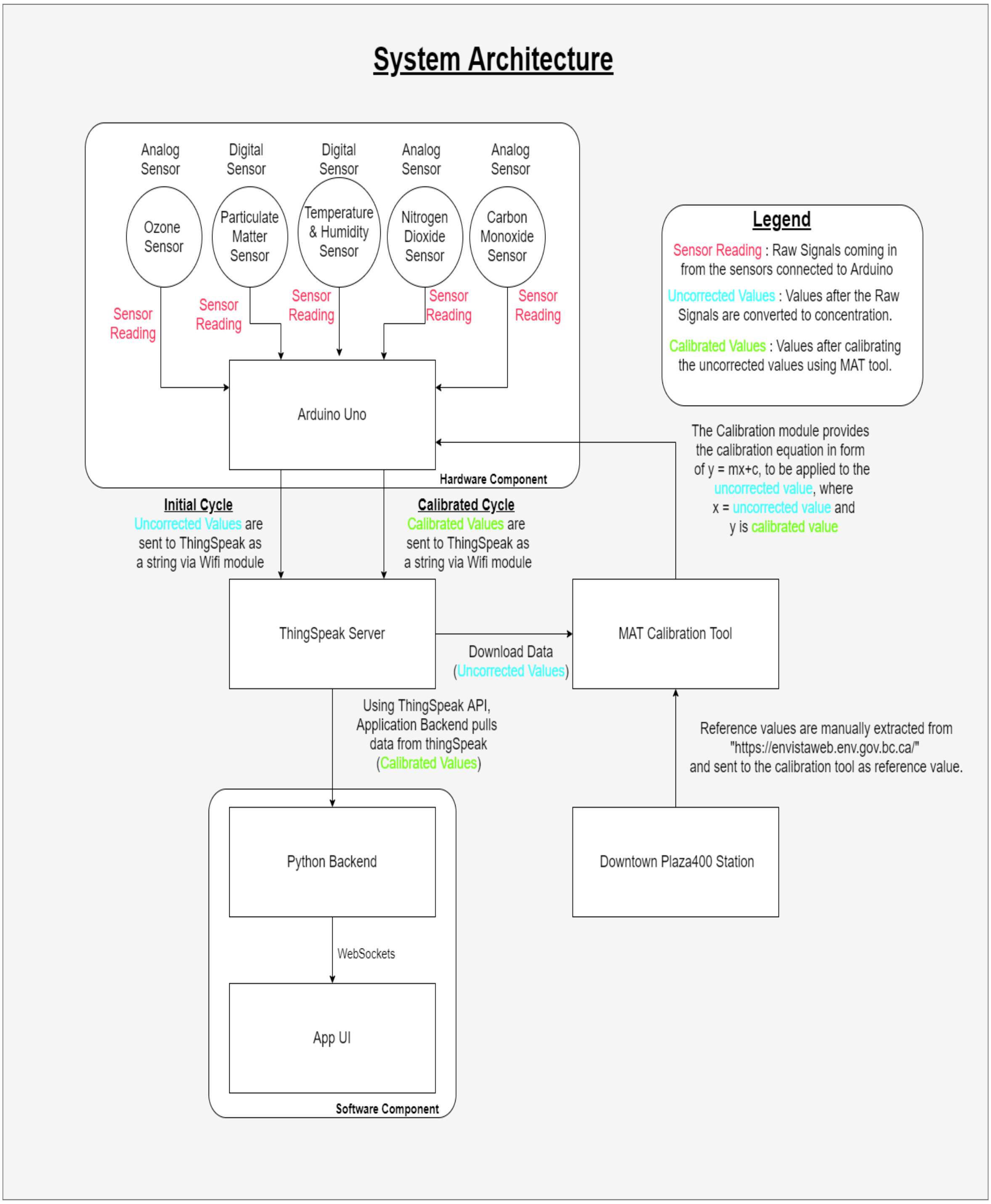


Figure 5.1: Flow diagram of Sensor system

5.1 Deployment

To understand the performance of our system, we deployed the air pollution monitoring system for eight days in University Heights, Prince George in which the first two days the raw data from the system was used for calibration and the following six days for measurement. Figure 5.2 shows the experimental set up of the sensor system deployed at the location. The system was directly connected to a power cable and was hung over a platform. The collected values were transferred to the ThingSpeak database through the WiFi module in an hourly averaged form and hence 24 data points were collected for each day.

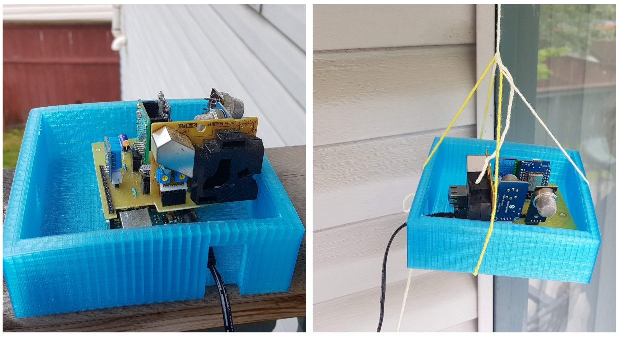


Figure 5.2: Deployed System at University Heights, Prince George

The main reason to conduct our deployment in this location was to have easy access and to check on the system regularly in case of any sensor faults or signal distortions. The measured raw values in the ThingSpeak database were compared to the reference values recorded at the downtown Plaza 400 building with the exception of carbon monoxide as it is not one of the pollutants measured at the Prince George reference station. Reference values for carbon monoxide were not available and hence were not analyzed. The Ozone, Particulate Matter (both $PM_{2.5}$ and PM_{10}), Nitrogen Dioxide, temperature and humidity data was collected and compared these to values at the Prince George reference station.

5.2 Calibration

The sensors that we are using to measure the pollutants are low-cost sensors and hence the accuracy of these sensors are expected to be less than those from the monitoring station operated by local or state agencies. To improve the reliability of the low-cost system we have used the MAT calibration procedure described in chapter 4. By applying the MAT procedure the final result is a calibration curve obtained by fitting the most appropriate equation to the set of data [93]. Each sensor gives raw data values corresponding to the presence of gases or pollutant concentration.

The calibration procedure we used for the system is to collect the raw data values for two days and then apply the MAT calibration procedure. This will generate a calibration equation that is applied to the system to get the calibrated data. The calibration curve along with the reference data and sensor data is as shown in Figure 5.3 for Ozone, Nitrogen Dioxide and Particulate Matter. To explain the calibration procedure we use Ozone as an example. Figure 5.3 (top) shows three line graphs which include the reference data from the monitoring station, our sensor raw data, and the resultant calibrated curve. The calibration curve was obtained by applying linear regression. From the data graph, it can be seen that the Ozone raw data from our sensor system provided systematically lower values than the reference system and after applying the calibration equation the data could be aligned to be quite close to that of the reference system.

Pollutant	Regression equation
O_3	$y = 1.42 * x + 0.04$
NO_2	$y = 1.49 * x + 1.36$
$PM_{2.5}$	$y = 1.31 * x + 0.27$
PM_{10}	$y = 1.59 * x + 0.47$

Table 5.1: Regression equation for each sensor

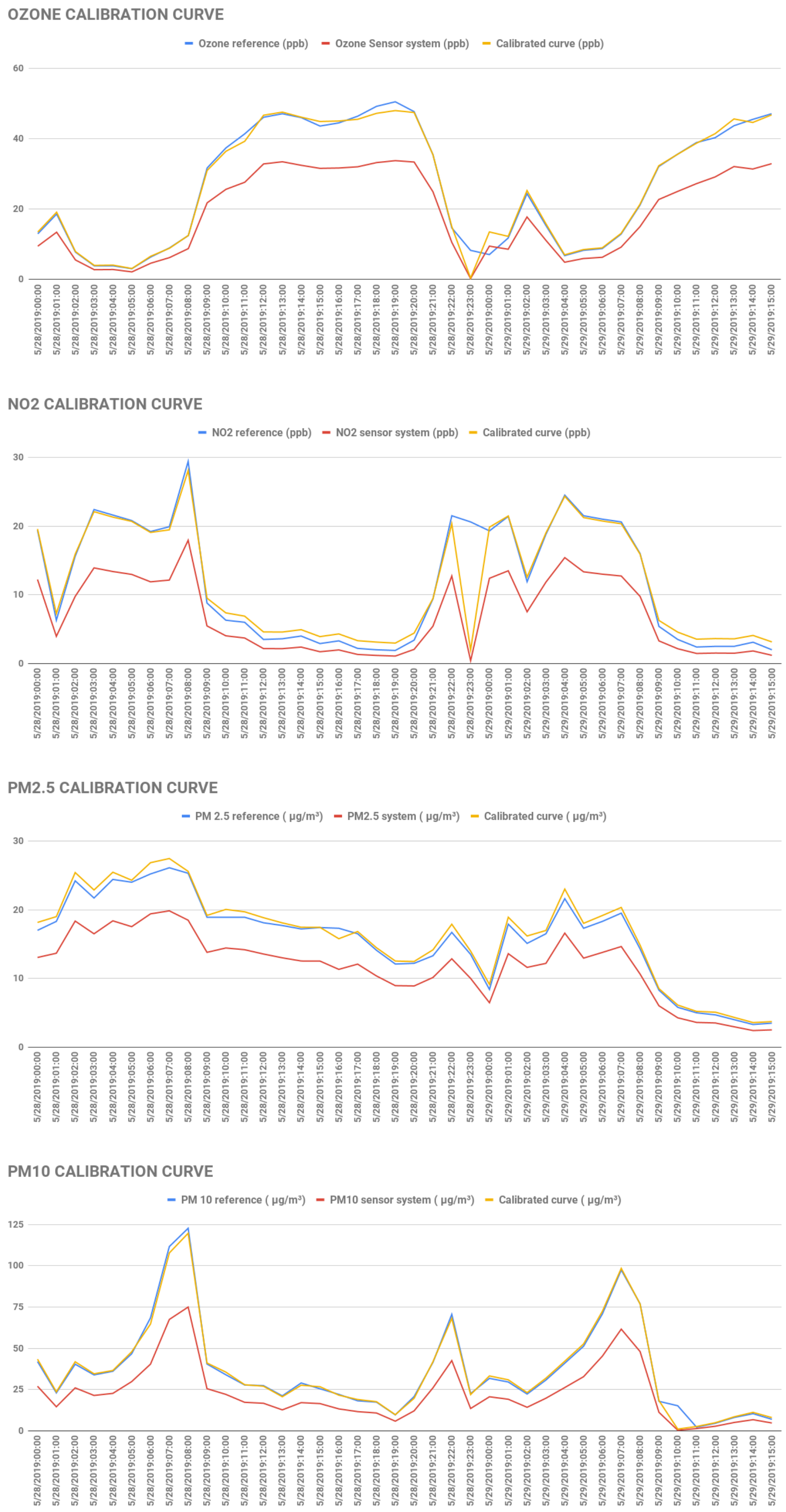


Figure 5.3: Calibration curves for Ozone, Nitrogen Dioxide, and Particulate Matter.

We calibrated all the sensors (except for carbon monoxide and temperature and humidity) with the same linear regression procedure using MAT. The calibration equation for our system is as shown in Table 5.1 where ‘x’ refers to the raw value and ‘y’ refers to the calibrated raw value. The calibration equations in Table5.1 were then supplied to the Arduino and then the system was used to measure the calibrated values.

5.3 Data Analysis

In this section, we will be comparing the collected data from our system to that from the reference system. We will be discussing the trends for each pollutant and will discuss possible reasons for the observed changes in concentration. The pollutants discussed are Ozone, Nitrogen Dioxide, $PM_{2.5}$ and PM_{10} and later we will also be discussing the aggregated pollutant indexes (AQI and AQHI) which we discussed in Chapter 1.

5.3.1 Ozone

As we have discussed in Chapter 1, Section 1.2, the Ground-Level Ozone (O_3) is a secondary pollutant and is not normally directly emitted into the atmosphere. It is produced from a complex series of chemical reactions between other pollutants like Nitrogen Oxides (NO_x) and volatile hydrocarbons (from combustion) in the presence of sunlight [5]. The main sources of Nitrogen Oxides and volatile hydrocarbons that results in the production of Ozone in Prince George are industrial facilities, motor vehicle exhaust, and chemical solvents. The provincial objective for Ozone is less than or equal to 63 ppb to meet the air quality standards [5].

Figures 5.4 and 5.5 shows the Ozone readings obtained each day from May 30 to June 4, 2019. From Figures 5.4 and 5.5 it can be seen that the calibrated sensor system shows a strong correlation with the reference system except for the last two days. The length of daylight on all the observed days was roughly around 16 hours and 30 minutes starting from

4.30 am in the morning to 9.30 pm in the evening [94]. In all the measured days it can be seen that the concentration of Ozone is high during the daytime when compared to the late evening and night hours. The several reasons for having a high Ozone level can be due to: [29]

1. " An increase in Ozone precursors (other pollutants like Carbon Monoxide, Methane, Nitrogen Oxide in the presence of sunlight helps in the production of Ozone) in the upper atmosphere"

2. "More snow cover at higher elevation causing higher reflection of solar radiation back to the atmosphere"

3. "High level of ozone in the stratosphere being transported to the troposphere"

4. "A combination of all the three factors"

In our case, the most likely factors are (1) and (3) or a combination of both. The possible reason for low concentration of Ozone during night hours is due to Ozone scavenging which occurs in the absence of sunlight [29]. Ozone scavenging is a chemical reaction between the Nitric Oxide with Ozone that is present in the atmosphere to form Nitrogen Dioxide (NO_2) [95]. Even though the pollutants like Nitrogen Dioxide and volatile hydrocarbons in the presence of sunlight contribute to the production of Ozone, Nitric Oxide can destroy it [95] [29].

Temperature is also considered to be a direct or indirect influence for an increase or decrease in the air pollution levels [29]. In the case of Ozone, higher temperature speeds up the reaction rate and increases the production of Ozone in the atmosphere [96]. For the measured days, the maximum temperature observed for the first three days (30th May to 1st June) is around 27 degrees Celsius and for the last three days (2nd June to 4th June) did not exceed 20 degrees Celsius. This might explain the observed higher concentrations of Ozone shown in Figure 5.4 when compared to the Figure in 5.5. The temperatures measured for each day by our calibrated system along with the comparison is included in the Appendix

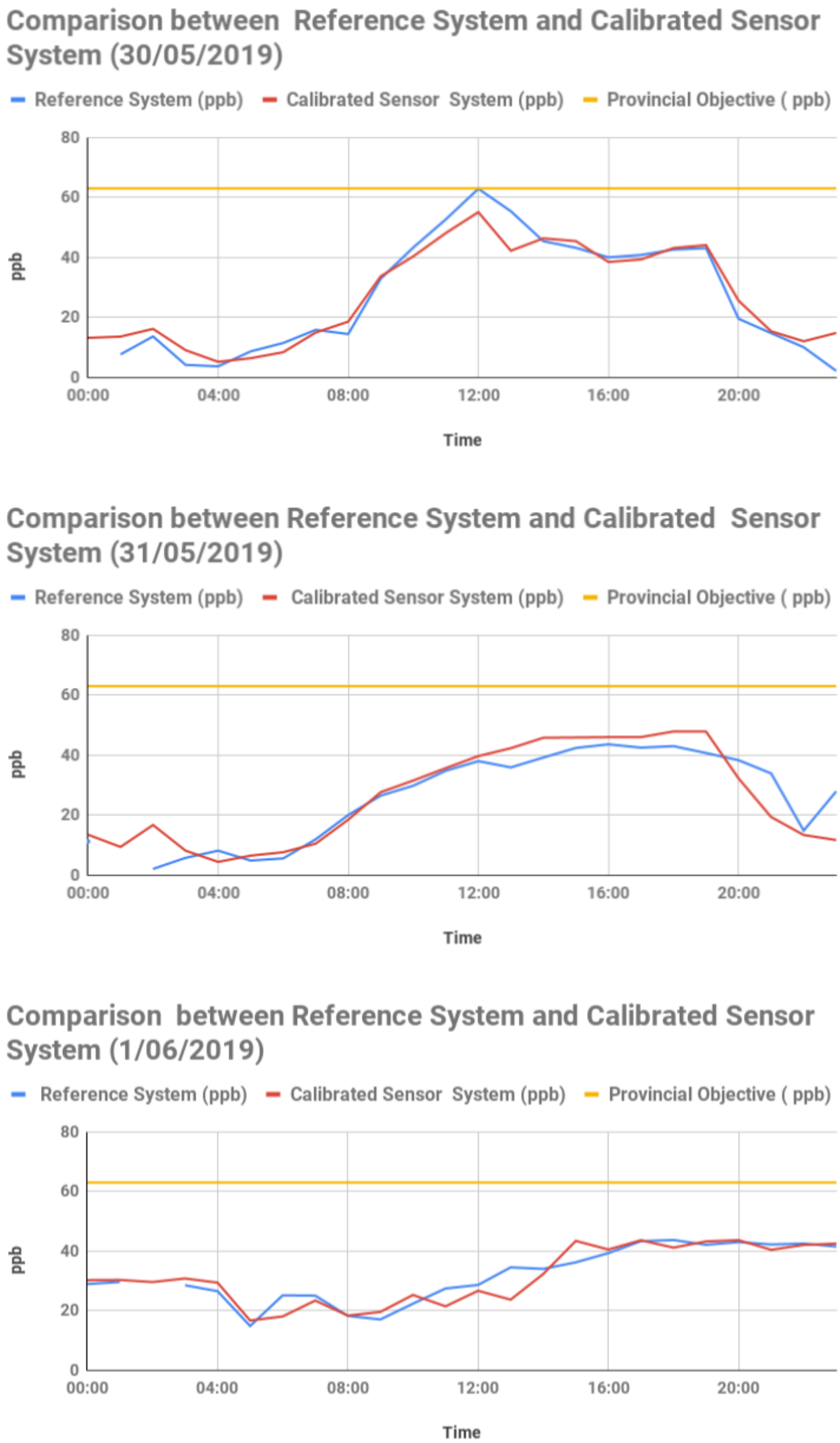


Figure 5.4: Comparison between Ozone values from sensor system and reference system from 30/05/2019 to 01/06/2019

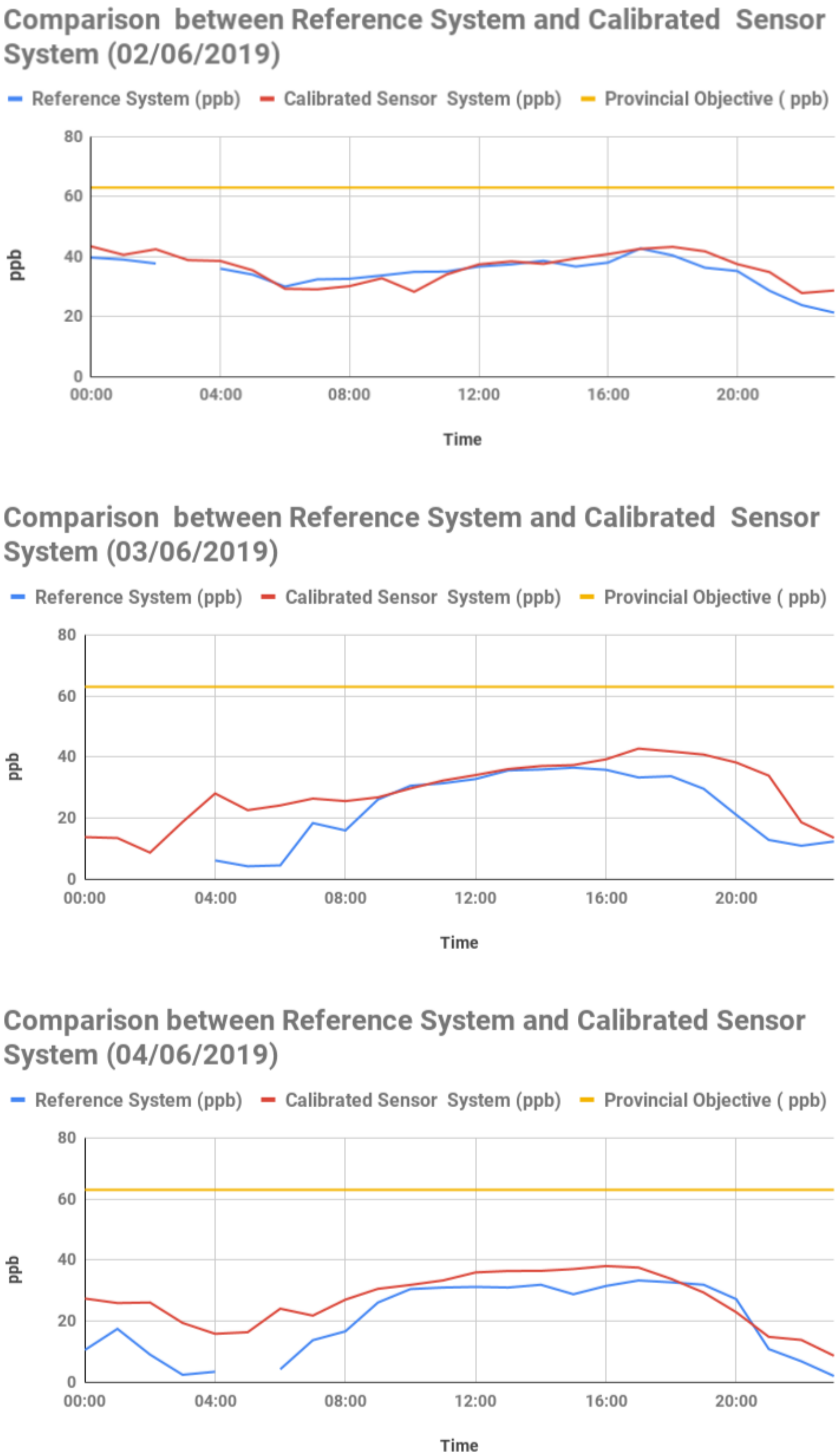


Figure 5.5: Comparison between Ozone values from sensor system and reference system from 02/06/2019 to 04/06/2019

section for reference. It can also be seen that the first three days it was sunny when compared to the last three days where it varied from partly sunny to rainy.

On all the measured days it can be seen that Ozone level have met the provincial objective standard which is less than or equal to 63 ppb [5]. Over the observed period the calibrated sensor system was able to capture the changes in the concentration of Ozone reliably. One outstanding issue is to determine when the system needs to do a recalibration procedure. As we have used a low-cost sensor system it is likely that there will be changes to the sensor performance over time. We have to characterize these sensor changes and the subsequent interval needs to be determined for recalibration procedures (eg: daily, weekly, quarterly). In general, we can say that the interval for recalibration for the sensor system needs to be looked in more detail. Nevertheless, it can also be seen that on the last two days the calibrated system showed differences in concentration when compared to the reference system. The environmental factors like temperature, relative humidity or concentration of confounding pollutants can result in change in sensor's sensitivity and reliability [97]. This can be a possible reason for the sensor response differences seen for the last two days.

5.3.2 Nitrogen Dioxide

Nitrogen Dioxide (NO_2) is one of the main pollutants in urban areas that penetrates deep into the lungs and can cause pulmonary diseases. It is mainly produced from high-temperature combustion of fossil fuels, power plants and industrial activities [29]. The main sources of NO_2 in Prince George include combustion of fossil fuels in vehicles, power plants, other industrial processes [5]. The provincial air quality objective for Nitrogen Dioxide is for concentrations to be less than to 32 ppb [5]. We have already discussed Nitrogen Dioxide in Chapter 1, section 1.2 about how high concentration of Nitrogen Dioxide can affect the human health. Using our system we have collected the data and have compared these to the data from the reference system. The collected sensor data in comparison with the reference system is shown in Figure 5.6 and 5.7 that covers the data collection from May 30 through

June 4, 2019.

From Figures 5.6 and 5.7, the observed trend in the concentration of Nitrogen Dioxide is that it is higher in the morning and evening, and has lower levels in the afternoon. The higher levels in the Nitrogen Dioxide in the morning and late evenings could be the result of the oxidation reaction between Nitric Oxide (NO) and Ozone (O_3) in the absence of sunlight. One hypothesis for the observed changes in concentration of Nitrogen Dioxide could be related to variation in solar radiation, since in the absence of the sunlight there will be more oxidation reaction between Nitric Oxide (NO) and Ozone (O_3) [79]. This process produces Nitrogen Dioxide as a night time reservoir and leads to increasing Ozone levels in the morning when photolysis resumes. This is the most likely reason for having a lower concentration of Nitrogen Dioxide during daytime [98]. Thus we could say that Ozone and Nitrogen Dioxide concentrations are often inversely related.

On all the measured days there were long hours of sunlight and short nights with the first three days having a clear sky and last three days with partly sunny and light rain. It can be seen from the Figure 5.6 and 5.7 that the concentration levels for Nitrogen Dioxide increases after 8:00 pm in the evening and decreases around 6:00 am in the morning and the reason for this could be related to the oxidation chemical reaction. The concentration of Nitrogen Dioxide for the last three days in Figure 5.7 is lower than the first three days in Figure 5.6. This can be related to the temperature and humidity as the temperature on the last three days did not exceed more than 20 degree Celsius and there were rain showers on 3rd June which can be the reason for a very low concentration of Nitrogen Dioxide during the daytime. During the first three days the temperature was around 27 degree Celsius and at that time the concentration of Nitrogen Dioxide was around 25 ppb during its peak hours except for June 1st where the concentration of Nitrogen Dioxide was around 12 ppb during its peak hours. This shows a positive correlation between temperature and concentration of Nitrogen Dioxide. The concentration of Nitrogen Dioxide did not exceed the provincial air quality objective of 32 ppb on any of the measured days but we do see that concentration was ranging from 10-20 ppb on most days.

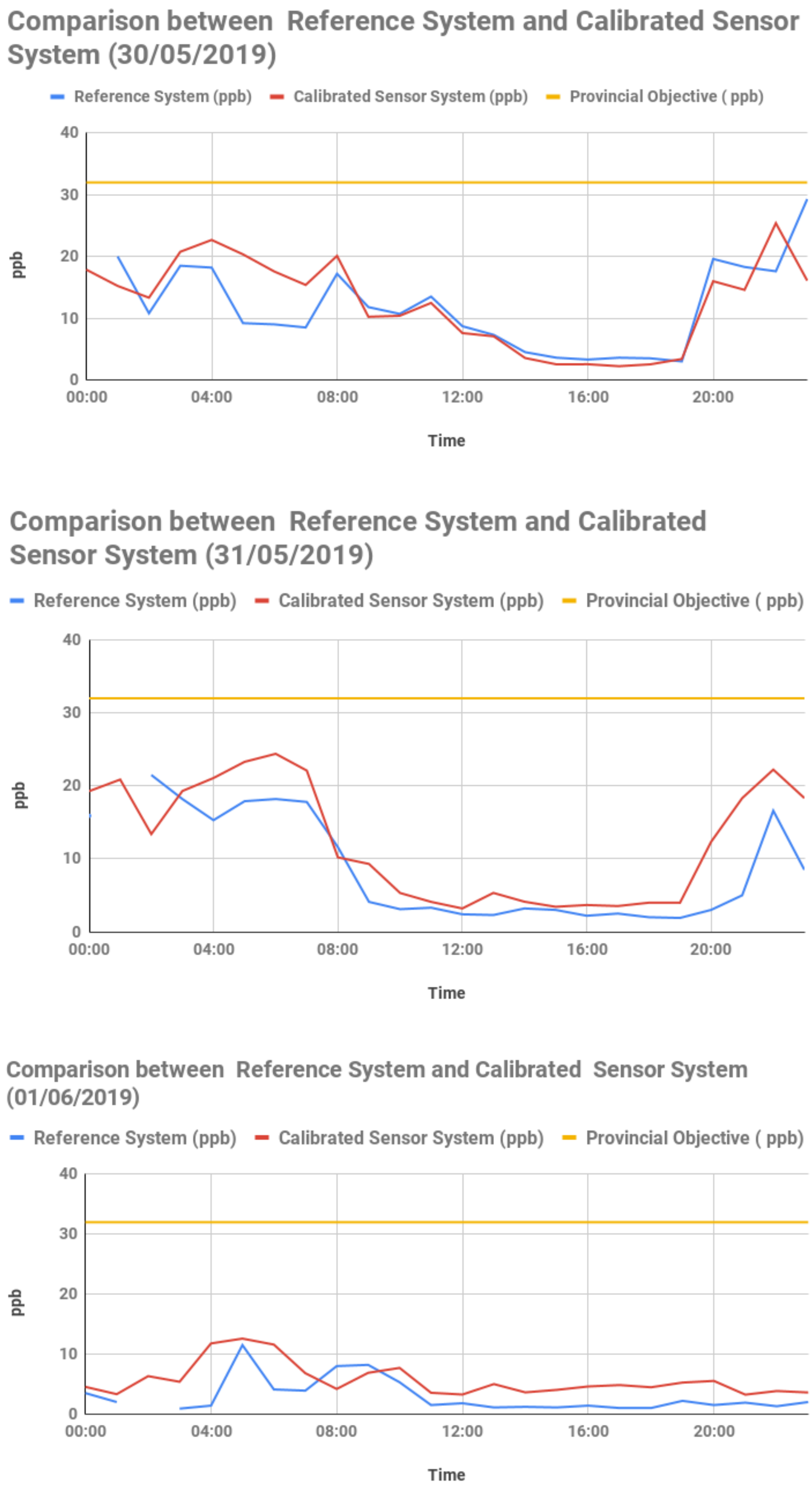


Figure 5.6: Comparison between Nitrogen Dioxide values from sensor system and reference system from 30/05/2019 to 01/06/2019

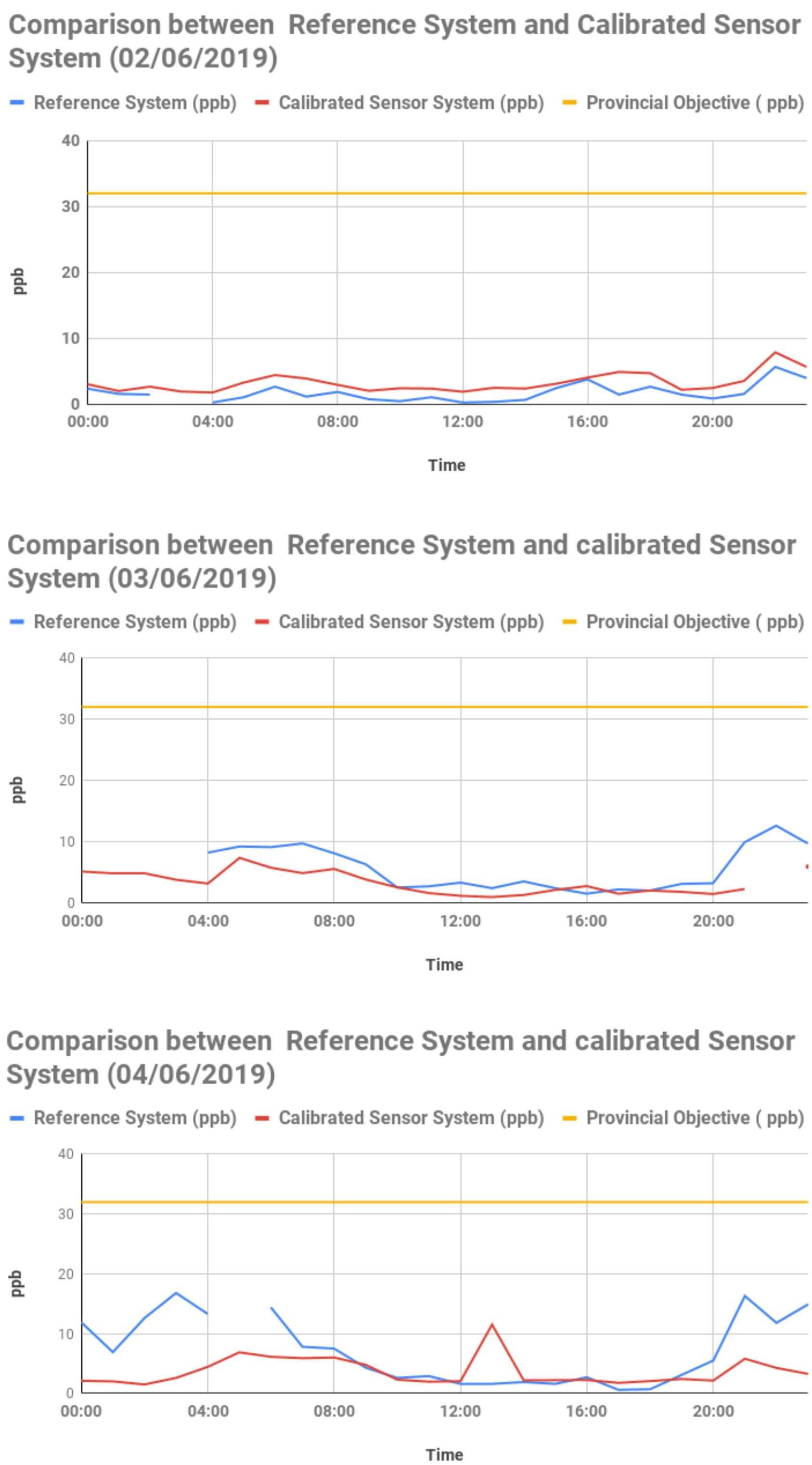


Figure 5.7: Comparison between Nitrogen Dioxide values from sensor system and reference system from 02/06/2019 to 04/06/2019

The results from our observations show that our NO_2 sensor had a strong correlation with the reference station for the first two days and later there was a fluctuation in the measurements. In absolute terms the differences are similar in all of the days and later days had less variability. There can be several possible reason for the variation in values. This can be either related to the drifting of sensor performance or the impact of environmental factors on the sensor. Another factor for less correlation can be the cross-sensitivity effect as all the sensors are placed together there will be the effect of other pollutant values which will cause a change in measurements [97].

Next factor could be change in temperature and relative humidity as these will reduce the sensitivity especially at higher humidity conditions leading to decrease in sensitivity [97]. Finally the location of the sensor system can also contribute to this as the system was placed in University Heights and the reference system is in Downtown.

5.3.3 Particulate Matter

Particulate Matter includes those particles whose size range from 0.001 µm to 100 µm. In Prince George, pulp mills and saw mills are primary sources for their emission, and during the winter, road dust also adds to this primary pollutant category as winter street sanding is done around that time [99]. The activities like street sweeping also contributes to the particulate matter pollution. Studies show that PM levels are usually low during the winter season and are higher during late winter and summer season [29]. In Prince George particulate matter pollution is a matter of serious concern and a variety of studies are being conducted. The provincial objective for $PM_{2.5}$ are 25 µg$/m^3$ and for PM_{10} are 50 µg$/m^3$ for the maximum daily (24-hours) mean for the year [5]. We will discuss the observed trends of the collected sensor data and the reference data from 30th May, 2019 to 4th June, 2019.

Both $PM_{2.5}$ and PM_{10} are location-dependent pollutants and can drastically change in concentration due to local activity. Data from the sensor system and the reference data from downtown for both $PM_{2.5}$ and PM_{10} are shown in Figures 5.8 to 5.11. Figures 5.8 and

5.9, presents PM_{10} data, and it can be seen that the values in the early morning and in the evening are the highest each day except for last two days in Figure 5.9. The daytime (11:00 am to 7:00 pm) concentration of PM_{10} is observed to be low generally. This is true in the case of the reference station values as well. On one of the measured days (on 03/06/2019), there was precipitation in the early morning from 12:00 am to 7:30 am and that might be the reason for lower concentrations of PM_{10} for that day and next (04/06/2019). The first two days showed a higher PM_{10} concentration compared to the rest of the measured days and can be be seen from the Figure 5.8 and 5.9. This higher concentration in PM_{10} can be associated to wind condition around that time. The first two days had calm winds around 4 km/h that makes it difficult to flush the PM_{10} pollutant out of the airshed and results in higher concentrations [29]. The rest of the days the wind speed varied from 10 to 20 km/hr and can be a reason for lower concentrations of PM_{10} when compared to the first two days. The provincial objective of PM_{10} (50 $\mu g/m^3$) was exceeded on the first three days.

In the case of $PM_{2.5}$, from the data in Figures 5.10 and 5.11 we can see similar trends from the sensor system data as the reference system except for some variations which happened for a short interval of time and these might be associated with local activities near the reference system or the sensor system. The first two days in Figure 5.10 showed a higher value for both sensor system and reference compared to the rest of the measured days. This can again be related to calm wind condition on for those two days when compared to the rest of the days.

In the Figures 5.8 to 5.11, it can be seen that for most of the days our sensor system measured higher values for both $PM2.5$ and $PM10$. This could be because the reference system is placed in a high roof top elevation and our sensor system is at a low roof top elevation. This could be one of the reason why our system shows higher value than the reference system. During the year 1996 a short-term experiment was conducted by Ministry of Environment to understand if there is any difference in the street-level pollution and was compared to the Plaza site (elevated building). The values collected from street levels shows that the particulate matter values were 30% higher than the Plaza value [79].

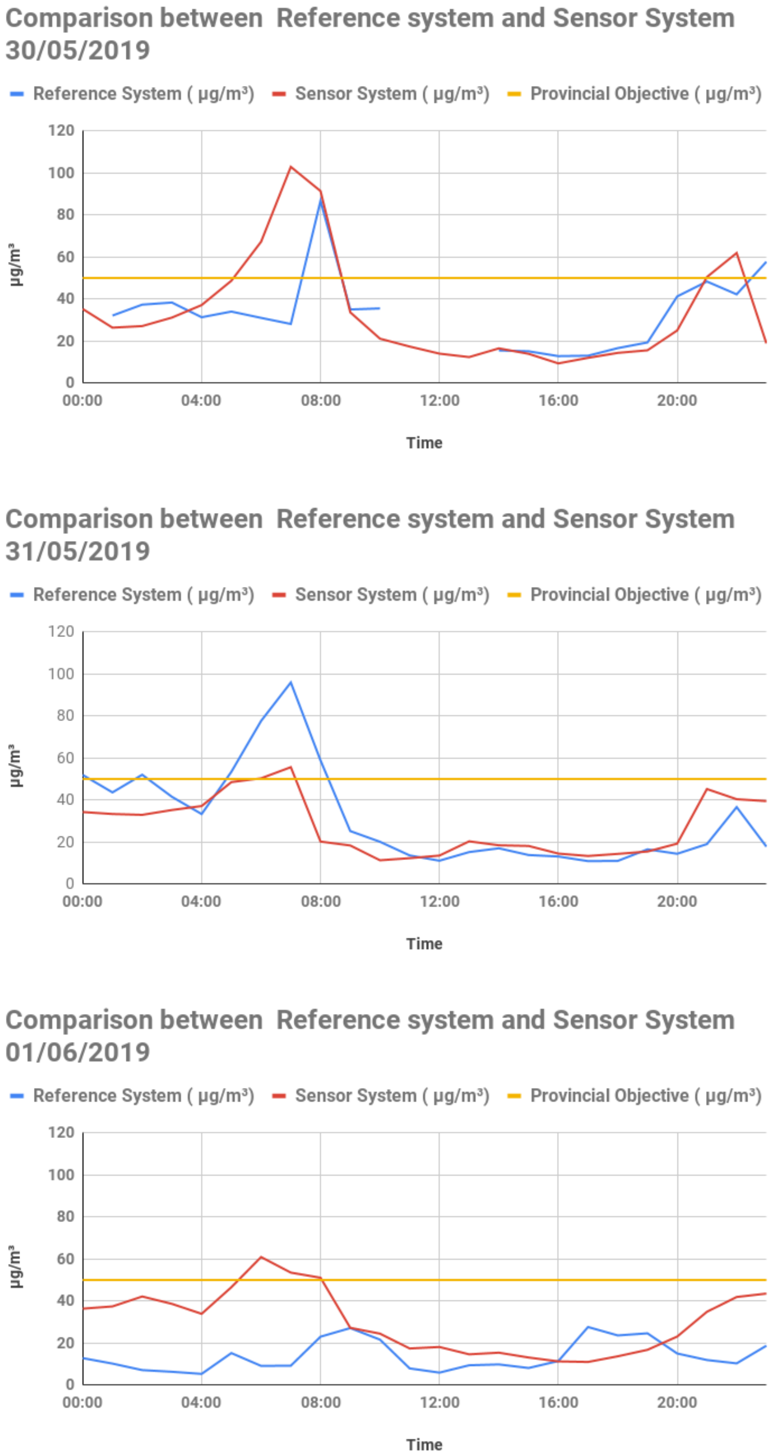


Figure 5.8: Comparison between PM_{10} values from sensor system and reference system from 30/05/2019 to 01/06/2019

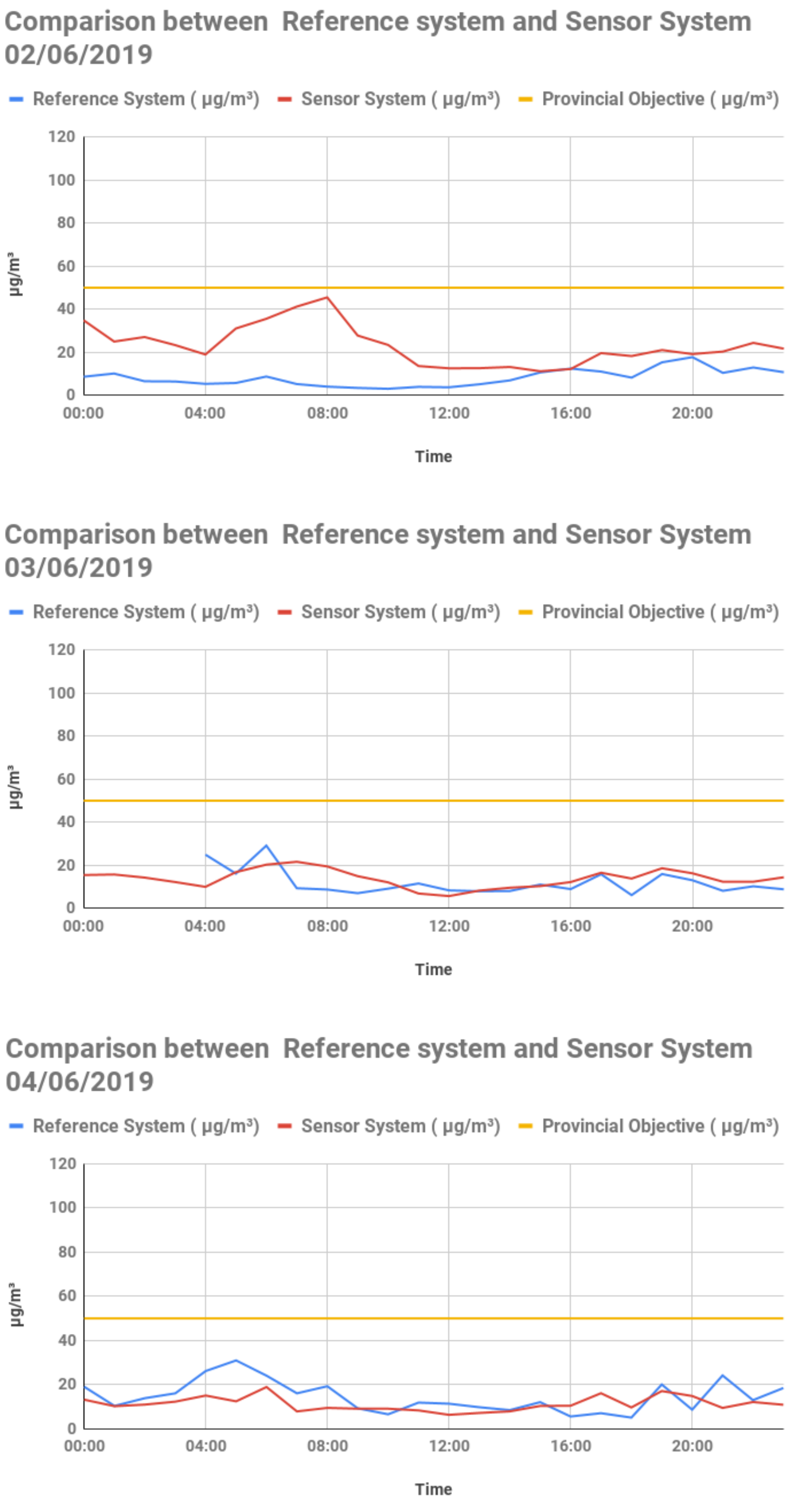


Figure 5.9: Comparison between PM_{10} values from sensor system and reference system from 02/05/2019 to 04/06/2019

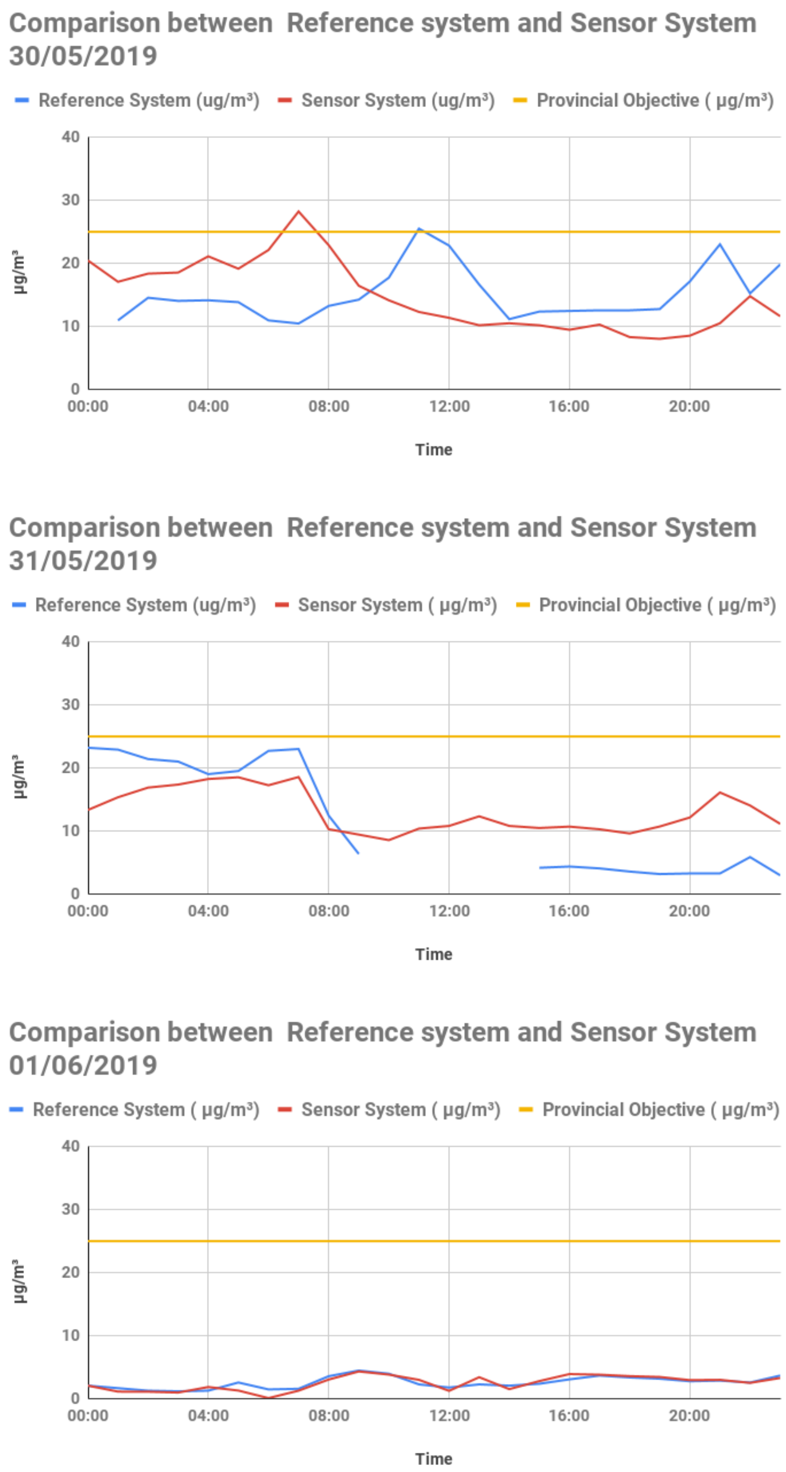


Figure 5.10: Comparison between $PM_{2.5}$ values from sensor system and reference system from 30/05/2019 to 01/06/2019

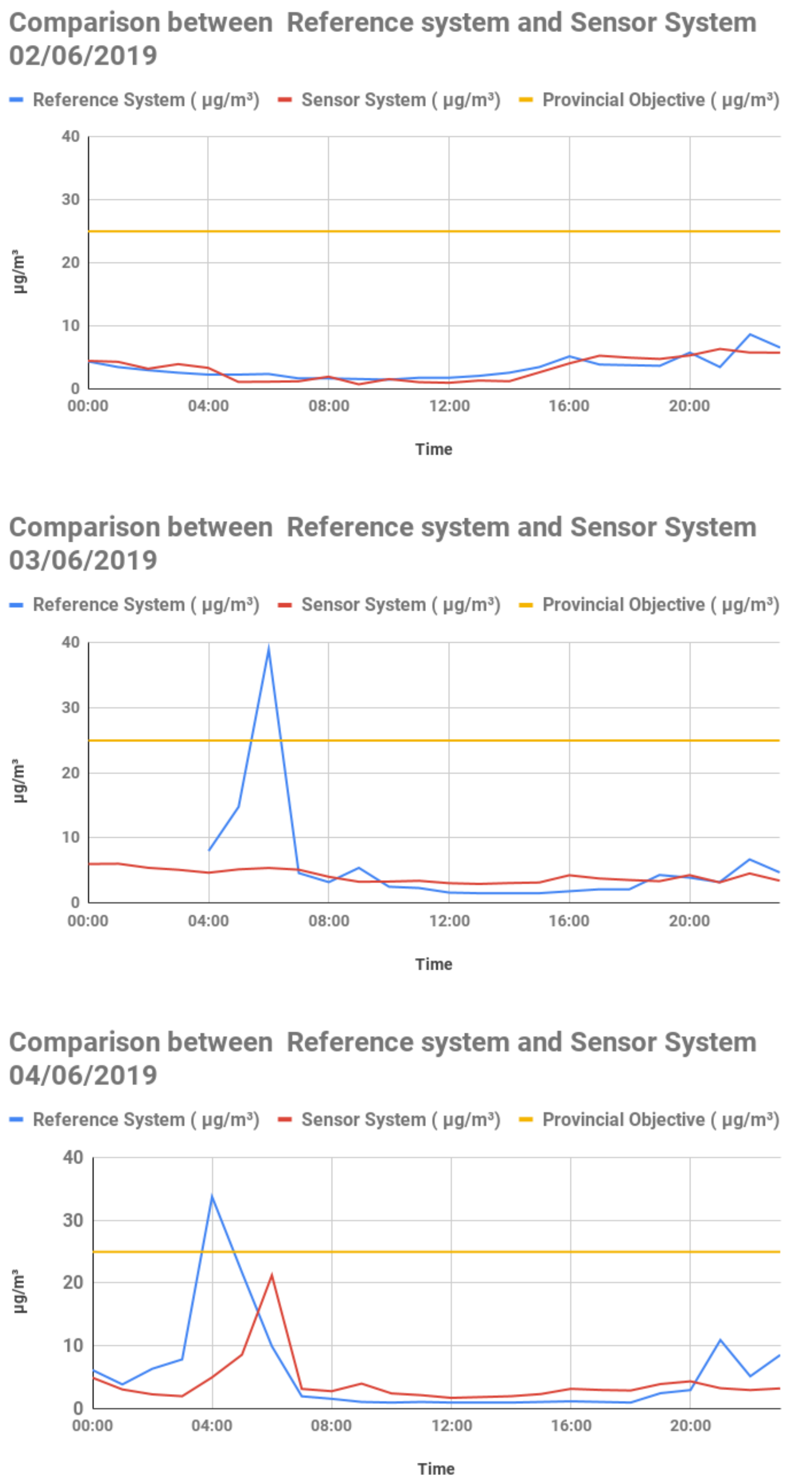


Figure 5.11: Comparison between $PM_{2.5}$ values from sensor system and reference system from 02/05/2019 to 04/06/2019

From this, we can say that as our sensor station is at a low roof top elevation could be a reason for showing higher concentration. Another possible factor for the difference in values in reference and sensor system can be the location where our sensor is placed and the wind direction. As our sensor was in a residential area on a roof top of a house (in College Heights, Prince George) the activities happening in the residential area could affect the concentration of Particulate Matter. The activities taking place in the downtown area are different to that of the residential area and the system could measure different values.

Both PM_{10} and $PM_{2.5}$ showed higher values for the first two days when compared to the rest of the days. Windspeed and precipitation are two factors that affect both PM_{10} and $PM_{2.5}$. High windspeed will contribute to dispersion of the particles into the atmosphere. The higher windspeed will help to transport the particles to a further region. Precipitation can also have a similar effect on Particulate Matter. Precipitation helps to settle down the dust particles and will result in lower concentrations.

In the Figure 5.9 and 5.11 we could see several peaks for both of the Particulate Matter pollutants that our sensor system measured. This is very tricky to identify, as unlike other pollutants present in the atmosphere both PM_{10} and $PM_{2.5}$ pollutant level are very dependent on location and any local activities has significant effect on the values.

5.3.4 Air Quality Indexes

As discussed in Chapter 1, section 1.2, the air quality indexes are the numbers that aid the general public in understanding the air quality in the area they live in. AQHI is a cumulative tool that helps to recognize the cumulative nature of poor air quality on health [37]. AQI is a single number that is calculated by selecting the maximum sub-indices of individual pollutants. In our analysis, we have calculated the AQI (Air Quality Index) and AQHI (Air Quality Health Index) indexes from the collected values of both our system and the reference system by using the equations 1.1, 1.2, and 1.3 from Chapter 1, section 1.2.

In Canada, the most commonly used index for the public is AQHI, which is a number scale which defines how the air quality affects the human health [100]. The value of AQHI ranges from 1 to 10+ (highest). The pollutants included in the calculation of this health index are NO_2, $PM_{2.5}$ and O_3. We have calculated the AQHI for each hour and then we took the average for the day, and have compared this with the AQHI provided by the Environment Canada based on the data from the reference system at Plaza 400. Figure 5.12 shows the comparison between these values. It can be seen from the Figure 5.12 that AQHI value is a combination of all the three pollutants. Ozone contributed the most to AQHI value in all the measured days for both sensor and reference system.

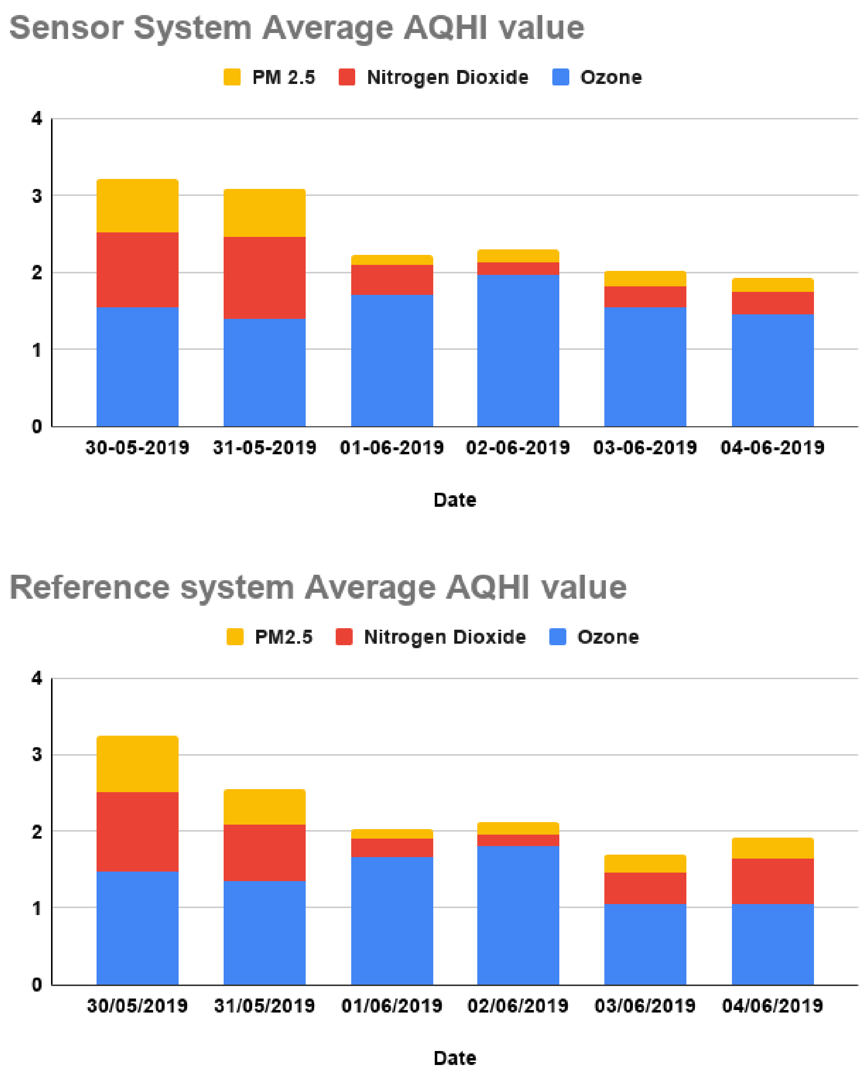


Figure 5.12: Comparison between average AQHI from both Reference and Sensor System

The overall AQHI for the observed days falls under the first category, which is the low-health risk which is discussed in Chapter 1, Section 1.2. From Figure 5.12 it can be seen that the AQHI values for all the days shows a good correlation to the reference system. On all the measured days the risk factor associated with health is low as per the index chart provided by the government [100] which is discussed in Chapter 1, Section 1.2.

The next index that we calculated is the AQI in which we have used the measured pollutants from the sensor system and the reference system. The pollutants used of measurement for the AQI are $NO2$, $PM2.5$ and $O3$ and we have not used CO as the reference system in Prince George did not provide the value. AQI communicates the air quality of an area based on the single worst pollutant. The Figure 5.13 shows the graph of the daily averaged AQI values.

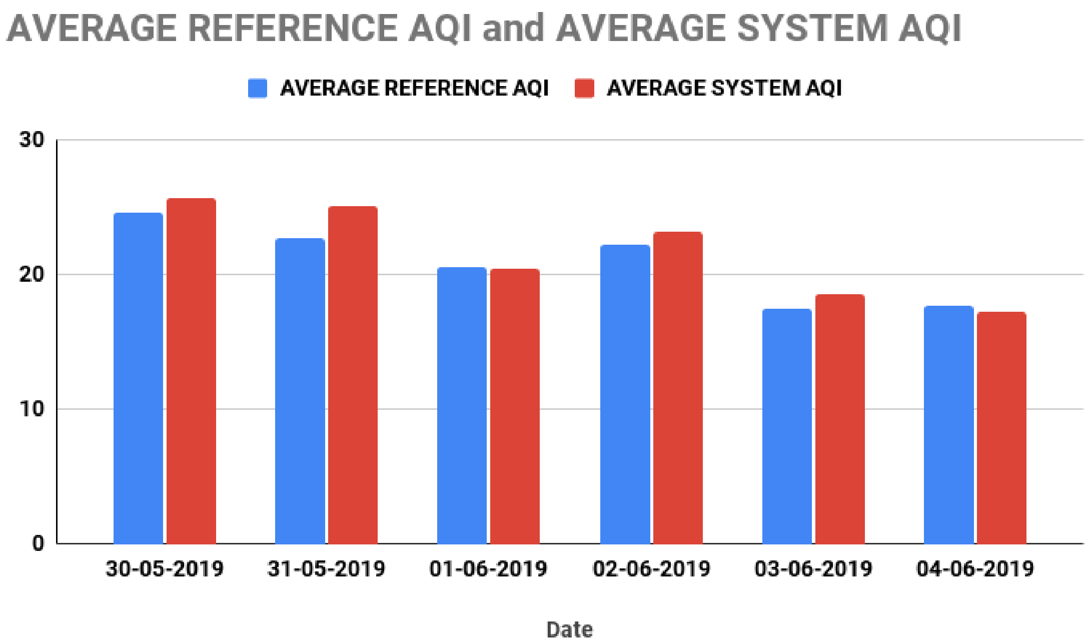


Figure 5.13: Comparison between average AQI from both Reference and Sensor System

It can be seen in Figure 5.13 that on majority of the days our sensor system shows a slightly higher AQI value as compared to the reference system. These AQI values fall into the first category of the scale provided by the government which means the air quality during the measured days was 'Good'. The pollutant which contributed the most to AQI is Ozone. BY looking at the correlation in reading of the AQHI and AQI index measurements in Figure 5.12 and 5.13 shows how well the system could capture the values after calibration.

In this work, we have built a sensor system which measures the pollutants specific to Prince George. The system which we have built is very simple, easily replicable and the sensors used are low-cost and readily available.We have also implemented calibration with the help of MAT tool and after that put our system out for six days and collected data. The measurement of each pollutant showed a good correlation to the reference system except for $PM_{2.5}$ and PM_{10}. The comparison of AQI and AQHI shows that our system was able to reproduce values similar to the reference system. One main concern of the system is the calibration procedures and the interval for recalibration. In order to understand this we need to collect more data. Overall we could say the system showed a good correlation to reference system and the calibration procedure could be improved. We will be discussing more about possible future work in the next chapter.

Chapter 6

Conclusion and Future work

The advancement in sensor technologies has given a lot of opportunities for researchers to explore different fields. These technical advancements have opened the door for solving major issues and one of them is environmental issues like air pollution. Air pollution has been a major concern ever since industrialization and modernization has hit the world and from then onwards it went on increasing. There are several works done under this category of research. In this chapter we will be summarizing the work done and also how this research project can be extended in future.

6.1 Summary

In our work, we have attempted to build a low-cost sensor system that measures the pollutants specific to the City of Prince George. The system which we built is simple, easily replicable and all the sensors used for the study are low-cost and readily available in the market. We have implemented calibration procedure to ensure the quality of measurement obtained from the system. This calibration procedure was a replication of MAT tool developed by Environmental Protection Agency (EPA). The calibrated data was visualized in an IoT platform for three categories of users which are layman, data scientists, and policy-makers. The system was deployed for six days for collecting the measurements and from the graphs, the measurements obtained from the system show correlation with the reference data obtained from Downtown Prince George.

6.2 Challenges and Directions Identified

Even though the system which we build showed a good response when compared to the reference system there are few downsides for the created system. One of the main drawbacks of the system was in the calibration procedure that we have applied. As discussed in chapter 4, we have replicated the MAT that uses linear regression algorithm for obtaining the calibration curve. From our observations, it can be seen that there are certain data points that are significantly different from the rest of the data points and these are called outliers. These outliers happen when there a variablility in the measured value or it could be a recording error from the sensor. In either cases we could visually plot the data to see the outliers in the measurements and understand from these points.

Next point of discussion is regarding the location of the sensor system for measurement. We have placed our sensor system in College Heights, Prince George that is around 9 km far from the reference system. Pollutant like Particulate Matter is very location-dependent and small activity could cause variation to the data. One solution for this issue could be possible field co-locations of the sensor to the reference system to understand on their performance [101]. This means that collecting the data from the same location to that of reference station will give more understanding.

Another matter of concern is regarding the unknown duration that the applied calibration is valid or to find out the recalibration interval so that the sensor's performance would not affect. As the sensor used for our work are low-cost sensors the problems may arise like drift in sensor's performance or faulty of sensor at deployment. To avoid these issues there should be routine checks on sensor's performance and also should determine the recalibration interval for the sensor. These interval could be different for different sensors and should be updated periodically as per the procedures in Chapter 4, section 4.2.

One of the main goal of our work was to build an efficient system with the help of low-cost sensors. The problem with the low-cost system is that there are variety of manufactures who produce these sensors. In our case we referred the Air Sensor Guidebook discussed in Chapter 3, section 3.1 for the selection of sensors. Even then the data sheet given by the manufactures lacks in detail, accuracy and reproducibility is unknown for most of the sensors. Another difficulty with low-cost sensor is regarding their sensitivity to changing environmental conditions mainly temperature, humidity, interference from other sensors, and signal drift. We also have not much knowledge of their long term stability and performance of the sensors as we only have measurements for six days.

6.3 Future Works

There are many ways this work could go forward and some of them I have listed below:

1. The system could be elaborated with different set of pollutants. In the present system we have only used a certain set of pollutant but there is a possibility to expand the system by adding another pollutant like TRS (Total Reduced Sulphur) or SO_2.

2. The sensors we have used for our studies will not work during extreme winter weather. There are sensors available in market which can replace the sensors used in our studies.

3. We could make the system better by applying collecting more data points which will give us a better understanding on the calibration procedure and the reliability of the sensors.

4. The calibration used here is linear regression which is an estimate of the 'best line' through the data. There will be some uncertainty associated to this and hence we can explore other calibration techniques.

5. Location where the instrument is placed is another main factor for measurement. By keeping the system as close as to the reference system will give a better idea on the

pollutant data.

6. We could work on the outlier data to understand if they have significant effect on the measurement. This could be done by filtering the outliers to understand the effects.

Bibliography

[1] Daniel C Mussatti, Margaret Groeber, and Dan Maloney. Generic Equipment and Devices for continuous emission monitoring system. *Evironment Protection Agency*, pages 1–40, 2000.

[2] World Health Organization. *Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. World Health Organization, 2006.

[3] Arnab Bhattacharya and Mukesh Sharma. National Air Quality Index. pages 1–58, 2014.

[4] British Columbia Government. The mobile air monitoring laboratory. `https://www2.gov.bc.ca/gov/content/environment/air-land-water/air/air-quality/measuring/maml`. Accessed: 2020-09-28.

[5] British Columbia Ministry of Environment. Air Quality in Prince George, Summary Report. pages 1–21, 2016. Accessed at: `https://princegeorge.ca/cityhall/mayorcouncil/councilagendasminutes/Agendas/2016/2016-09-19/documents/Attch_AQ_prince-george-aq-report-june2016.pdf`.

[6] Shu Chiung Hu, You Chiun Wang, Chiuan Yu Huang, and Yu Chee Tseng. A vehicular wireless sensor network for CO2monitoring. *Proceedings of IEEE Sensors*, pages 1498–1501, 2009.

[7] Rundong Tian, Christine Dierk, Christopher Myers, and Eric Paulos. MyPart: Personal, Portable, Accurate Airborne Particle Counting. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1338–1348, 2016.

[8] P Dutta, Paul M Aoki, N Kumar, and A Mainwaring. Common Sense: Participatory urban sensing using a network of handheld air quality monitors. *SenSys '09*, pages 349–350, 2009.

[9] Alyssa I. Robinson Fletcher, Richard Ribón Nicolas M. Oreskovic, M.D. Design and Clinical Feasibility of Personal Wearable Monitor for Measurement of Activity and Environmental Exposure. pages 874–877, 2015.

[10] B. Fang, Q. Xu, T. Park, and M. Zhang. AirSense: An intelligent home-based sensing system for indoor air quality analytics. *UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 109–119, 2016.

[11] National Exposure Research Laboratory Office of Research and Development. How to Evaluate Low-Cost Sensors by Collocation with Federal Reference Method Monitors. pages 1–45, 2017.

[12] Jacobs Technology Inc, Microsoft Corporation and EPA. Macro analysis tool. `https://www.epa.gov/sites/production/files/2018-01/macro_analysis_tool.xlsm`, September 2016.

[13] The free Encyclopedia Wikipedia. Linear regression. `https://en.wikipedia.org/wiki/Linear_regression`, 2018. [Online; accessed 04-December-2018].

[14] Aaron Daly and Paolo Zannetti. An introduction to air pollution–definitions, classifications, and history. *Ambient air pollution. P. Zannetti, D. Al-Ajmi and S. Al-Rashied, The Arab School for Science and Technology and The EnviroComp Institute*, pages 1–14, 2007.

[15] Ioannis Manisalidis, Elisavet Stavropoulou, Agathangelos Stavropoulos, and Eugenia Bezirtzoglou. Environmental and health impacts of air pollution: A review. *Frontiers in public health*, 8, 2020.

[16] Claes Bernes. Persistent organic pollutants. *Swedish Environmental Protection Agency, Stockholm*, pages 9–21, 1998.

[17] L Ritter, KR Solomon, J Forget, M Stemeroff, and C O'Leary. Persistent organic pollutants: an assessment report on: Ddt, aldrin, dieldrin, endrin, chlordane, heptachlor, hexachlorobenzene, mirex, toxaphene, polychlorinated biphenyls, dioxins and furans. *International Programme on Chemical Safety (IPCS)*, 1995.

[18] Health Effects Institute. State of Global Air 2017 : A Special Report. *Boston, MA:Health Effects Institute.*, page 24, 2017.

[19] A. Wolman. Public health and environment. *American Journal of Public Health*, 75(9):1049–1051, 1985.

[20] KC Heidorn. A chronology of important events in the history of air pollution meteorology to 1970. *Bulletin of the American Meteorological Society*, 59(12):1589–1597, 1978.

[21] Viktor P Ál. The Environmental Consequences of Industrialization in Western European Core Countries and the Borsod Basin of Hungary , 1850 – 1945 : A Comparative Outline. 3:75–84, 2016.

[22] ET Wilkins. Air pollution and the london fog of december, 1952. *Journal of the Royal Sanitary Institute*, 74(1):1–21, 1954.

[23] K Balakrishnan, M Brauer, G Chen, and J Chow. *To Humans Outdoor Air Pollution*, volume 109. 2015.

[24] Daniel Vallero. Tool Steel - an overview (pdf) | ScienceDirect Topics. *Fundamentals of Air Pollution (Fifth Edition)*, pages 1–10, 2014.

[25] William E. Wilson and Helen H. Suh. Fine particles and coarse particles: Concentration relationships relevant to epidemiologic studies. *Journal of the Air and Waste Management Association*, 47(12):1238–1249, 1997.

[26] C.M. Payus, A.T. Vasu Thevan, and J. Sentian. Impact of school traffic on outdoor carbon monoxide levels. *City and Environment Interactions*, 4(2019):100032, 2019.

[27] M Artha P Atricia Sierra-vargas and L U I S M Teran. INVITED REVIEW SERIES : AIR POLLUTION AND LUNG HEALTH Air pollution : Impact and prevention. (January), 2012.

[28] Farideh Golbabaei, Hamid Hassani, Asghar Ghahri, and Mirghani Seyed Someah. Risk Assessment of Exposure to Gases Released by Welding Processes in Iranian Natural Gas Transmission Pipelines Industry. 4(2):6–9, 2012.

[29] Environmental Quality Section, Environmental Protection Division, Ministry of Environment. 2010 Annual Air Quality Report Prince George Airshed. pages 1–126, June, 2012. Accessed at : `https://www2.gov.bc.ca/assets/gov/environment/air-land-water/air/reports-pub/2010-aq-report.pdf`.

[30] Heidi Salonen, Tunga Salthammer, and Lidia Morawska. Human exposure to NO2 in school and office indoor environments. *Environment International*, 130(May):104887, 2019.

[31] United states Environmental Protection Agency. Ground-level ozone basics. `https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics#wwh`, 2018.

[32] Morton Lippmann. Health Effects Of Ozone A Critical Review. *Journal of the Air Pollution Control Association*, 39(5):672–695, 1989.

[33] Kirsten Cukor. Bridging the divide between air quality monitoring, management and policy in the sea-to-sky airshed: A method for analyzing and interpreting large volume air quality data for management and policy guidance. *IIIEE Master thesis*, 2015.

[34] Hong Chen and Ray Copes. *Review of Air Quality Index and Air Quality Health Index: Environmental and Occupation Health*. Public Health Ontario, 2013.

[35] Wikipedia. Air quality index. `https://en.wikipedia.org/wiki/Air_quality_index`. Accessed: 2019-09-30.

[36] Nikhil Soni and Neetishree Soni. Benefits of pedestrianization and warrants to pedestrianize an area. *Land Use Policy*, 57:139–150, 2016.

[37] Paul Hasselback and Eric Taylor. *Air quality health index variation across British Columbia*. Interior Health, 2010. Accessed at: `https://www2.gov.bc.ca/assets/gov/environment/air-land-water/air/reports-pub/aqhi-variation-bc.pdf`.

[38] British Columbia Government. Measuring air quality. https://www2.gov.bc.ca/gov/content/environment/air-land-water/air/air-quality/measuring. Accessed: 2020-09-28.

[39] Pollution Prevention Ministry of Water Land and Air Protection Omineca-Peace Region. 2000 annual air quality report. pages 1–144, April, 2002.

[40] Northern Health Authority and Ministry of Environment. Prince George Mobile Monitoring Final Report. pages 1–77, April, 2011.

[41] Government of British Columbia. Prince george plaza 400 - air monitoring station. http://www.env.gov.bc.ca/epd/bcairquality/data/station.html?id=0450307, 2020.

[42] Gary Anthes. Open source software no longer optional. *Communications of the ACM*, 59(8):15–17, 2016.

[43] Kavi Kumar Khedo and Vishwakarma Chikhooreeah. Low-Cost Energy-Efficient Air Quality Monitoring System Using Wireless Sensor Network. *Wireless Sensor Networks - Insights and Innovations*, (October), 2017.

[44] Yang Huang Liu, Liang Hu, Disheng Yang, and Hengchang. Air-Sense: indoor environment monitoring evaluation system based on ZigBee network. *IOP Conference Series: Earth and Environmental Science*, 81(1):12208, 2017.

[45] Wei Yi, Kin Lo, Terrence Mak, Kwong Leung, Yee Leung, and Mei Meng. *A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems*, volume 15. 2015.

[46] Movva Pavani and P. Trinatha Rao. Urban air pollution monitoring using wireless sensor networks: A comprehensive review. *International Journal of Communication Networks and Information Security*, 9(3):439–449, 2017.

[47] Anthony Downs. Traffic: What its getting worse, What government can do. (January), 2004.

[48] 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.

[49] Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. Real-time Air Quality Monitoring Through Mobile Sensing in Metropolitan Areas. 2013.

[50] Yoshinari Shirai, Yasue Kishino, Futoshi Naya, and Yutaka Yanagisawa. Toward On-Demand Urban Air Quality Monitoring using Public Vehicles. *IWSC@Middleware*, pages 1:1–1:6, 2016.

[51] Chaosheng Xiang, Xiaofeng Liu, Aimin Jiang, Bin Yan, and Jing Xia. Poster: A Bicycle-borne Sensor Network for Monitoring Urban Air Quality. *Mobisys'16: Companion Companion Publication of the 14Th Annual International Conference on Mobile Systems, Applications, and Services*, (March):88, 2016.

[52] Xiaofeng Liu, Bin Li, Aimin Jiang, Shixin Qi, Chaosheng Xiang, and Ning Xu. A bicycle-borne sensor for monitoring air pollution near roadways. *2015 IEEE International Conference on Consumer Electronics - Taiwan, ICCE-TW 2015*, (January 2017):166–167, 2015.

[53] S. D. Zhi, Y. B. Wei, and Z. H. Yu. Air Quality Monitoring Platform based on Remote Unmanned Aerial Vehicle with Wireless Communication. *Proceedings of the International Conference on Future Networks and Distributed Systems - ICFNDS '17*, (July):1–7, 2017.

[54] P M Aoki, R J Honicky, A Mainwaring, C Myers, E Paulos, S Subramanian, and A Woodruff. Common sense: Mobile environmental sensing platforms to support community action and citizen science. *Adjunct Proceedings Ubicomp 2008*, (October):Paper 201 OR – Human Computer Interaction Institu, 2008.

[55] Piero Zappi, Elizabeth Bales, Jing Hong Park, William Griswold, and Tajana Šimunić Rosing. The CitiSense Air Quality Monitoring Mobile Sensor Node. *Proceedings of the 11th ACM/IEEE Conference on Information Processing in Sensor Networks, Beijing, China*, (January), 2012.

[56] brian p rubin. Tzoa's Wearable Enviro-Tracker Wants To Clear The Air. `https://readwrite.com/2015/05/20/tzoa-wearable-air-quality-sensor-crowdfunding-indiegogo/`, 2015. [Online; accessed 06-March-2019].

[57] Chris C. Lim Michael Heimbinder, Raymond Yap. The aircasting platform. `http://aircasting.org/`, 2006. [Online; accessed 01-January-2019].

[58] Dan Han. Air Quality Monitoring , Fushun-Kokkola. pages 1–2, 2010.

[59] Aircasting Luminescence, New York Hall, The Aircasting Luminescence, Iem Heng, Raymond Yap, City Tech, Andy Zhang, The Aircasting, and Luminescent Vest. Make Your Own AirCasting Luminescent Apparel.

[60] Alfredo J. Perez, Miguel A. Labrador, and Sean J. Barbeau. G-Sense: A scalable architecture for global sensing and monitoring. *IEEE Network*, 24(4):57–64, 2010.

[61] Diego Mendez, Alfredo J. Perez, Miguel A. Labrador, and Juan Jose Marron. P-Sense: A participatory sensing system for air pollution monitoring and control. *2011 IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2011*, pages 344–347, 2011.

[62] Man Wong, Tsan Yip, and Esmond Mok. Development of a Personal Integrated Environmental Monitoring System. *Sensors*, 14(11):22065–22081, 2014.

[63] Kavi K. Khedo, Rajiv Perseedoss, and Avinash Mungur. A Wireless Sensor Network Air Pollution Monitoring System. *International Journal of Wireless & Mobile Networks*, 2(2):31–45, 2010.

[64] M. F.M. Firdhous, B. H. Sudantha, and P. M. Karunaratne. IoT enabled proactive indoor air quality monitoring system for sustainable health management. *Proceedings of the 2017 2nd International Conference on Computing and Communications Technologies, ICCCT 2017*, (February):216–221, 2017.

[65] Rohan Narayana Murty, Geoffrey Mainland, Ian Rose, Atanu Roy Chowdhury, Abhimanyu Gosain, Josh Bers, and Matt Welsh. CitySense: An urban-scale wireless sensor network and testbed. *2008 IEEE International Conference on Technologies for Homeland Security, HST'08*, pages 583–588, 2008.

[66] John Bicket, Daniel Aguayo, Sanjit Biswas, and Robert Morris. Architecture and evaluation of an unplanned 802.11b mesh network. page 31, 2005.

[67] Joseph Camp, Joshua Robinson, Christopher Steger, and Edward Knightly. Measurement driven deployment of a two-tier urban mesh access network. page 96, 2006.

[68] CUWin Foundation. open source community hosting that makes a difference. `https://www.cuwireless.net/`, 2006. [Online; accessed 10-April-2019].

[69] Jen Hao Liu, Yu Fan Chen, Tzu Shiang Lin, Da Wei Lai, Tzai Hung Wen, Chih Hong Sun, Jehn Yih Juang, and Joe Air Jiang. Developed urban air quality monitoring system based on wireless sensor networks. *Proceedings of the International Conference on Sensing Technology, ICST*, pages 549–554, 2011.

[70] NATIONAL INSTRUMENTS. LabVIEW. (June):1–89, 2013.

[71] Aaron Daly and Paolo Zannetti. An Introduction to Air Pollution – Definitions , Classifications , and History. pages 1–14, 2007.

[72] Emily G. Snyder, Timothy H. Watkins, Paul A. Solomon, Eben D. Thoma, Ronald W. Williams, Gayle S. W. Hagler, David Shelow, David A. Hindin, Vasu J. Kilaru, and Peter W. Preuss. The Changing Paradigm of Air Pollution Monitoring. *Environmental Science & Technology*, 47(20):11369–11377, 2013.

[73] U.S. Environmental Protection Agency. Macro Analysis Tool -MAT. `https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=340520&Lab=NERL`, 2018. [Online; accessed 04-December-2018].

[74] Massimo Banzi and First Edition. *Getting started with Arduino*. :Books, an imprint of Maker Media, a division of O'Reilly Media, Inc., october 20 edition, 2008.

[75] Manuel Ruiz Guti. Arduino + Ethernet Shield. pages 1–42, 2017.

[76] Technical Data. MQ-2 Semiconductor Sensor for Alcohol. *Structure*, pages 3–5, 2012.

[77] Vaibhav Jain. Insight - Learn the Working of a Gas Sensor. `https://www.engineersgarage.com/insight/how-gas-sensor-works`, 2013. [Online; accessed 10-November-2018].

[78] SGX Sensortech. SGX Sensortech Data Sheet MiCS-2714 MOS Sensor for Nitrogen Dioxide. pages 1–5.

[79] Ministry of Environment. 2008 Annual Air Quality Report for the Prince George Airshed. pages 1–122, 2010.

[80] Tracy Allen. De-construction of the Shinyei PPD42NS dust sensor. (510):0–3, 2002.

[81] Joel Kuula, Timo Mäkelä, Risto Hillamo, and Hilkka Timonen. Response characterization of an inexpensive aerosol sensor. *Sensors (Switzerland)*, 17(12), 2017.

[82] Espressif Systems. ESP8266EX. pages 1–22, 2018.

[83] Mary R English. Who Are the Stakeholders in Environmental Risk Decisions? ... How Should They Be Involved? *Risk: Health, Safety & Environment*, 11(243):352–362, 2000.

[84] Katherine Hepworth. Big Data Visualization : Promises & Pitfalls. pages 7–19.

[85] Alberto Cairo. Ethical Infographics. *Investigative Reporters and Editors, Inc. The IRE Journal*, 37(2):25–27, 2014.

[86] ThingSpeak. ThingSpeak- IOT platform. `https://thingspeak.com/`, 2017. [Online; accessed 26-January-2018].

[87] Georgia Miskell, Jennifer A Salmond, and David E Williams. Solution to the problem of calibration of low-cost air quality measurement sensors in networks. *ACS sensors*, 3(4):832–843, 2018.

[88] Ron Williams, Tim Watkins, and Russell Long. Low cost sensor calibration options the straw-man approach, 2013. `https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=517654`.

[89] U.S. Environmental Protection Agency. Air Sensor Toolbox for Citizen Scientists, Researchers and Developers. `https://www.epa.gov/air-sensor-toolbox`, 2018. [Online; accessed 04-December-2018].

[90] Instruction guide and macro analysis tool for community-led air monitoring. `https://www.epa.gov/air-research/instruction-guide-and-macro-analysis-tool-community-led-air-monitoring`, 2016. Accessed: 2018-12-06.

[91] Ron Williams. Emerging Sensor Technologies. `https://www.frames.gov/documents/smoke/videos/SmoC_webinar_20180308_CharacterizingAirQualityInARapidlyChangingWorld_Williams-slides.pdf`, 2018. [Online; accessed 10-November-2018].

[92] U.S. Environmental Protection Agency. Calibration Curves: Program Use / Needs. https://www.epa.gov/sites/production/files/2014-05/documents/calibration-guide-ref-final-oct2010.pdf, 2010. [Online; accessed 04-December-2018].

[93] D Stone and J Ellis. Calibration and Linear Regression Analysis: A Self-Guided Tutorial (Part 2). pages 1–7, 2001.

[94] Prince George, British Columbia, Canada — sunrise, sunset, and daylength. https://www.timeanddate.com/sun/canada/prince-george?month=6&year=2019. Accessed: 2021-01-24.

[95] Ministry of Environment Ontario. Archived - Air Quality in Ontario 2016 Report. pages 1–39, June, 2019. accessed at https://www.ontario.ca/document/air-quality-ontario-2016-report/ground-level-ozone#:~:text=Ground%2Dlevel%20ozone%20is%20a,in%20the%20presence%20of%20sunlight.&text=Nitric%20oxide%20reacts%20quickly%20with,atoms%20(%20O%20)%20and%20NO%20.

[96] Jane Coates, Kathleen A Mar, Narendra Ojha, and Tim M Butler. The influence of temperature on ozone production under varying nox conditions–a modelling study. *Atmos. Chem. Phys*, 16(18):11601–11615, 2016.

[97] Andrea L Clements, William G Griswold, Abhijit Rs, Jill E Johnston, Megan M Herting, Jacob Thorson, Ashley Collier-Oxandale, and Michael Hannigan. Low-cost air quality monitoring tools: from research to practice (a workshop summary). *Sensors*, 17(11):2478, 2017.

[98] Katarzyna Rozbicka and Tomasz Rozbicki. Spatiotemporal variations of tropospheric ozone concentrations in the Warsaw agglomeration (Poland). *Annals of Warsaw University of Life Sciences-SGGW. Land Reclamation*, 46(3), 2014.

[99] Gary Champagne and Dennis Fudge. Prince George Air Quality Management Background Report. 1996. Accessed at : https://www.pgairquality.com/uploads/files/pdf/PRAQMB-Report(1997).pdf.

[100] Government of Canada. About the air quality health index. https://www.canada.ca/en/environment-climate-change/services/air-quality-health-index/about.html#What_is_AQHI, 2019.

[101] 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):e0137789, 2015.

[102] Shaun Burke and High Wycombe. Regression and Calibration. pages 13–18, 2001.

[103] S.Miller. The Method of Least Squares and Signal Analysis. pages 1–7, 1992.

[104] A. M. Almeida, M. M. Castel-Branco, and A. C. Falcão. Linear regression for calibration lines revisited: Weighting schemes for bioanalytical methods. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 774(2):215–222, 2002.

[105] Wayne W Daniel and Chad Lee Cross. *Biostatistics: a foundation for analysis in the health sciences*. Wiley, 2018.

[106] Time and date. Past Weather in Prince George, British Columbia, Canada — June 2019. https://www.timeanddate.com/weather/canada/prince-george/historic?month=6&year=2019, 2019. [Online; accessed 24-February-2021].

Appendix A

Appendix

A.1 Linear Regression

Calibration can be done through different statistical methods and the most popular approach is Linear Regression method. In statistics, the term regression is used to describe a group of methods that summarize the degree of association between one variable (or set of variables) and another variable (or set of variables) [102]. Linear Regression establishes relationship between two variables one is the independent variable which will be on the x axis and the dependent variable in the y axis by drawing a straight line. If there are many observations and is plotted as a scatter plot then a line called as regression line as in the figure A.1 could be drawn through it by the Least Square method.

The method of Least Squares is a procedure to determine the best fit line to data [103] by minimizing the errors between the observed value and the actual value. This method considers that, for each value of x, there is a sub-population of y values normally distributed, that the means of all the sub-populations of y lie on the same straight line and all the sub-populations of y values have equal variances [104] [105]. The line drawn will have a slope m and is given by the formula A.1

$$m = \frac{\sum_i [(x_i - \bar{x})(y_i - \bar{y})]}{[\sum_i (x_i - \bar{x})^2]} \tag{A.1}$$

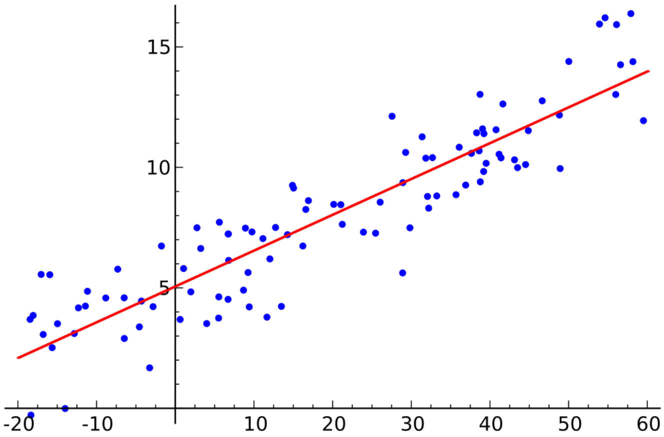


Figure A.1: Linear Regression Model [13]

where m is the slope, x_i and y_i are the reference and observed data values, $\bar{x}$ and $\bar{y}$ are the mean values of the data. The mean values will act as the centroid of the regression line and it has to pass through these points. The intercept equation c is

$$C = \bar{y} - m\bar{x} \tag{A.2}$$

On getting the slope and intercept value the equation of line can be found. Once the Regression line is drawn, to find out how much the data is scattered and to what extend, correlation coefficient (R^2) is used. In other words R^2 gives the measure of degree to which the values of x and y are linearly correlated [93]. The equation for finding the regression value is

$$R^2 = \frac{\sum_i[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{[\sum_i(x_i - \bar{x})^2[\sum_i(y_i - \bar{y})^2]]}} \tag{A.3}$$

The range of value for R^2 is between 0 to 1 and closer the value is to 1, the stronger the correlation between the two data. All these equation can be easily calculated with the help of Excel and thus all these equations are integrated to the MAT tool.

A.2 Weather in Prince George

In this section we have included the weather data that includes the temperature, humidity and the sunlight data in Prince George for the measured days. We have shown a comparison of the collected data from the temperature and humidity sensor of the sensor system to the reference system in downtown Prince George. The data of the sunlight was obtained from the weather website which stores the previous weather data of the city [106].

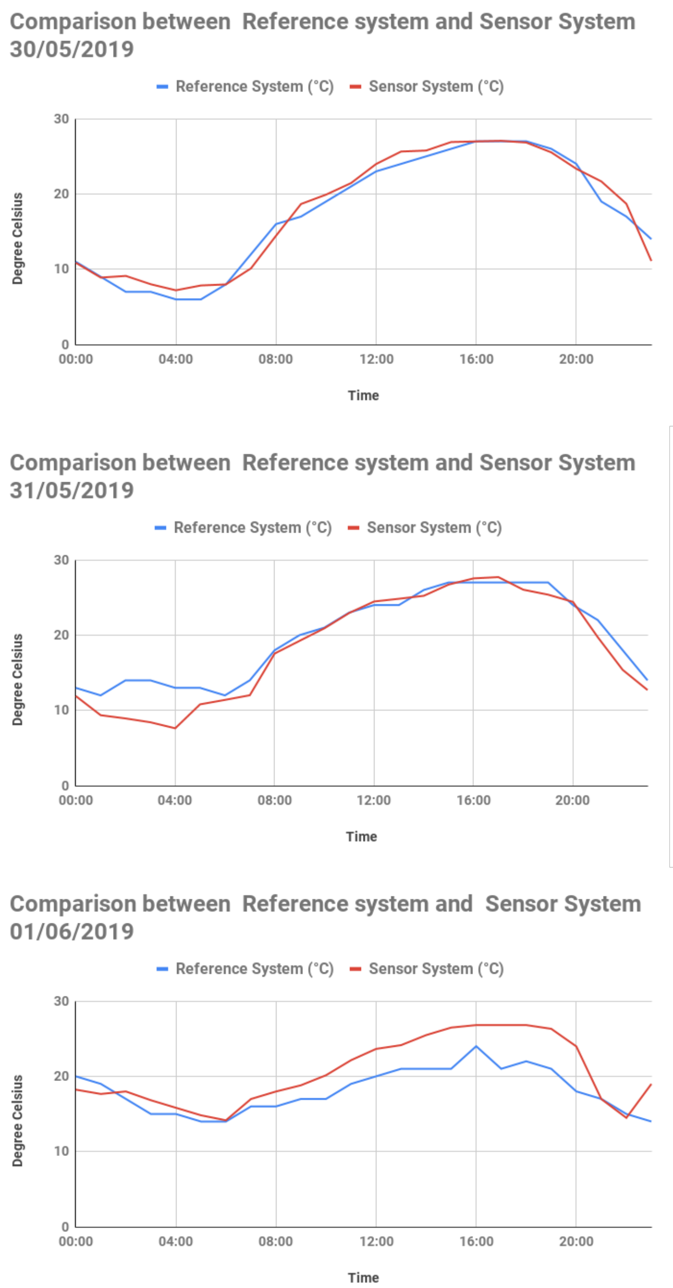


Figure A.2: Comparison between Temperature values from the sensor system and reference system from 30/05/2019 to 01/06/2019

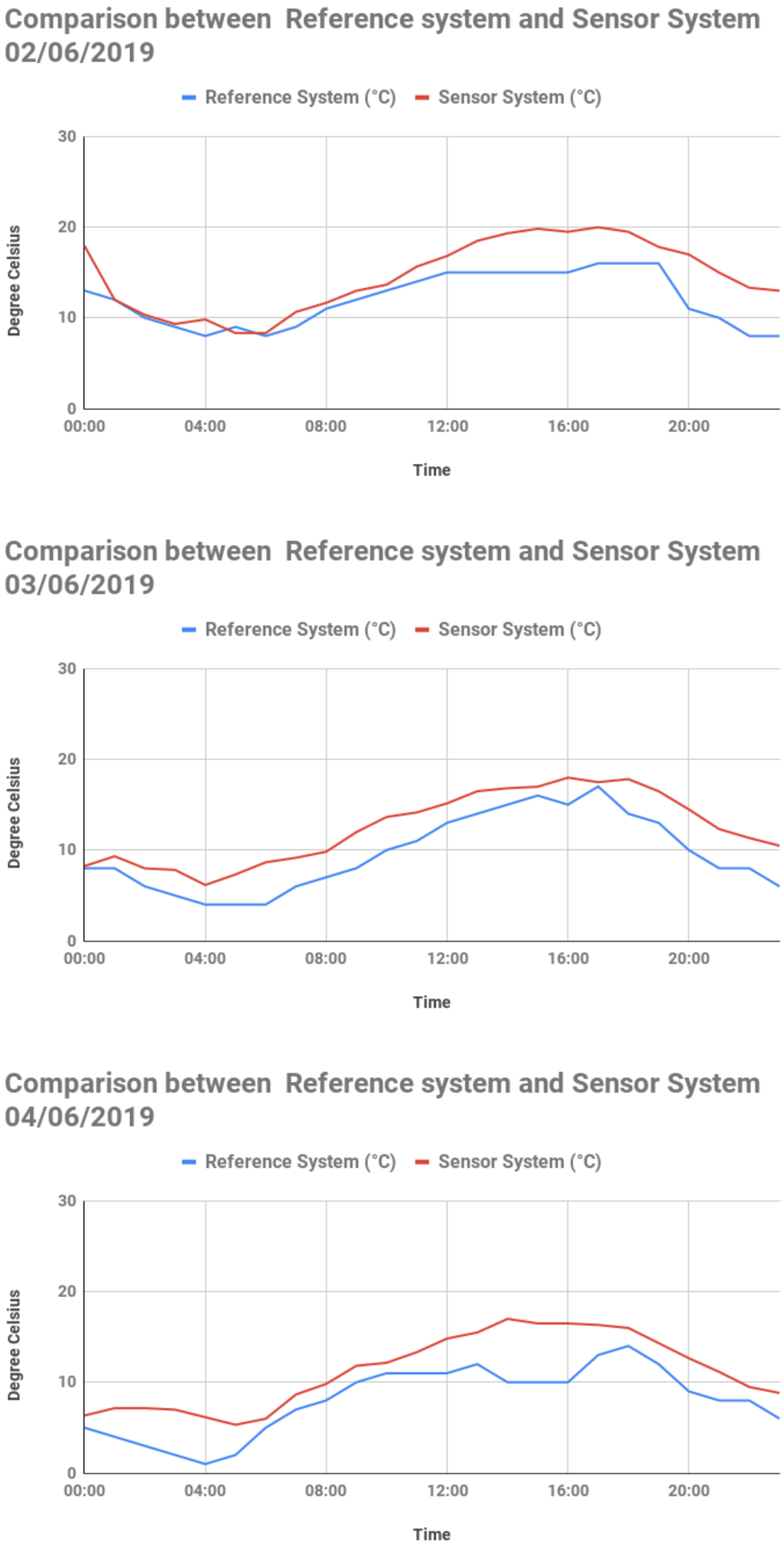


Figure A.3: Comparison between Temperature values from the sensor system and reference system from 02/06/2019 to 04/06/2019

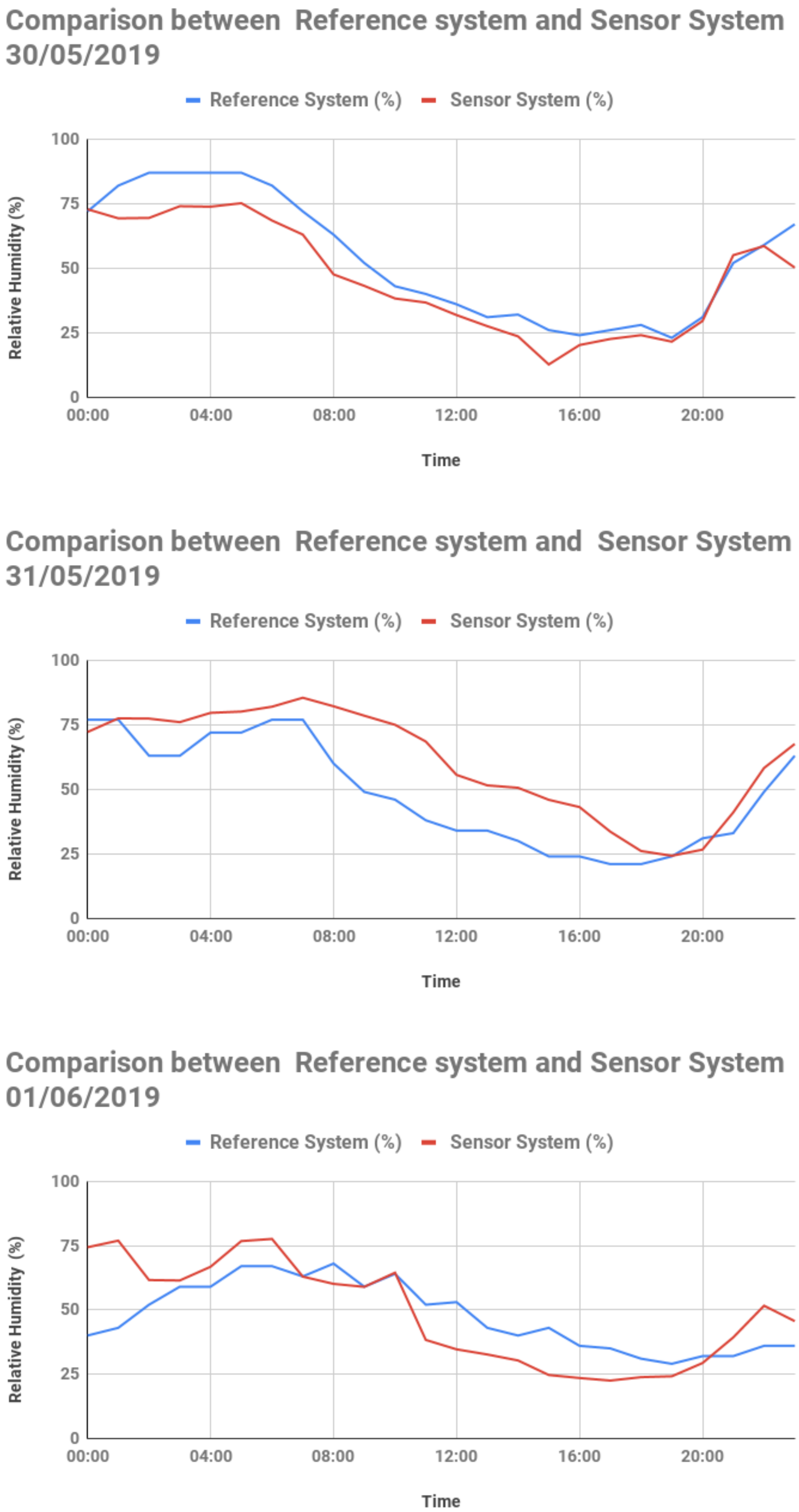


Figure A.4: Comparison between Humidity values from the sensor system and reference system from 30/05/2019 to 01/06/2019

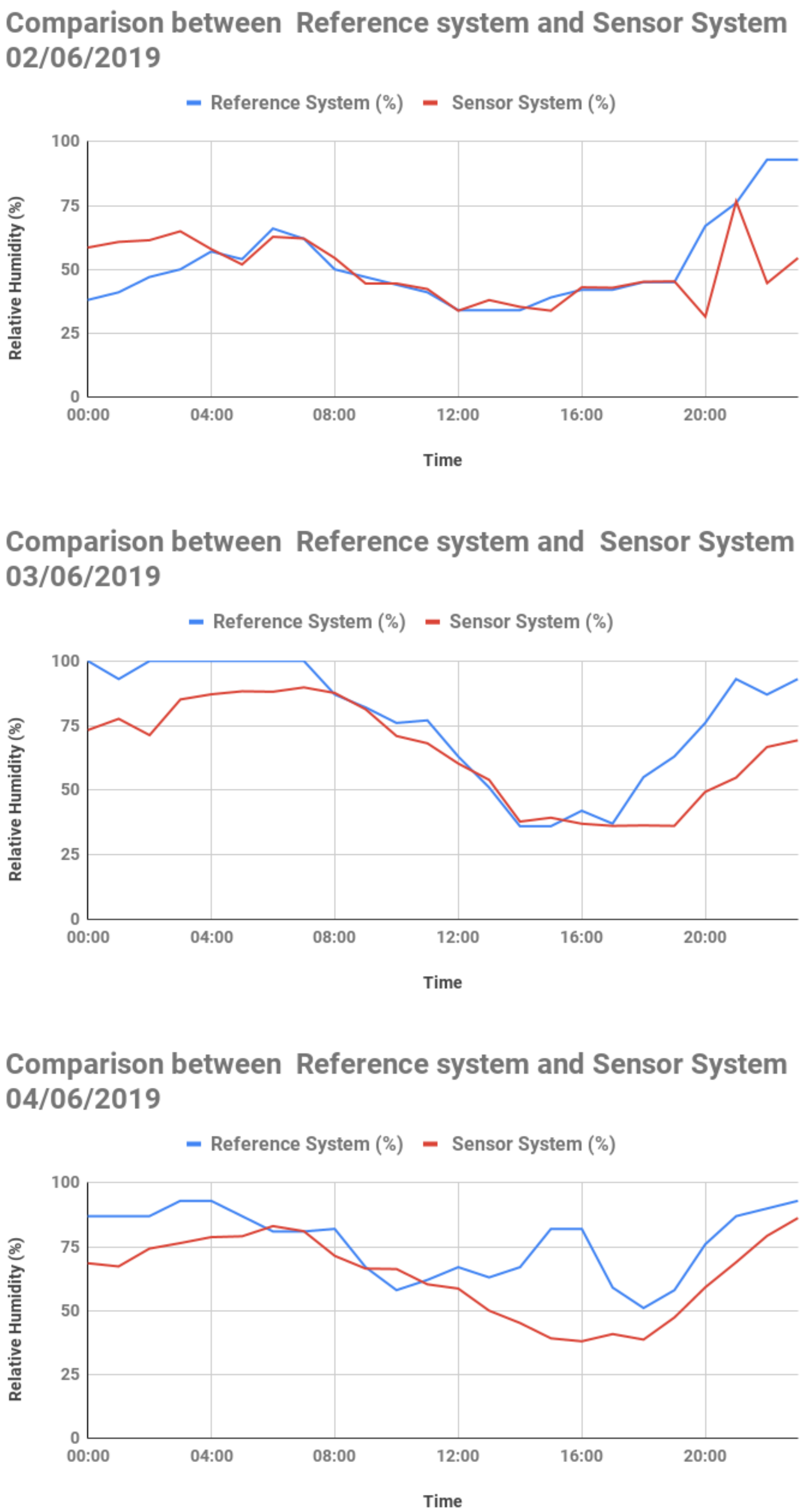


Figure A.5: Comparison between Humidity values from the sensor system and reference system from 02/06/2019 to 04/06/2019

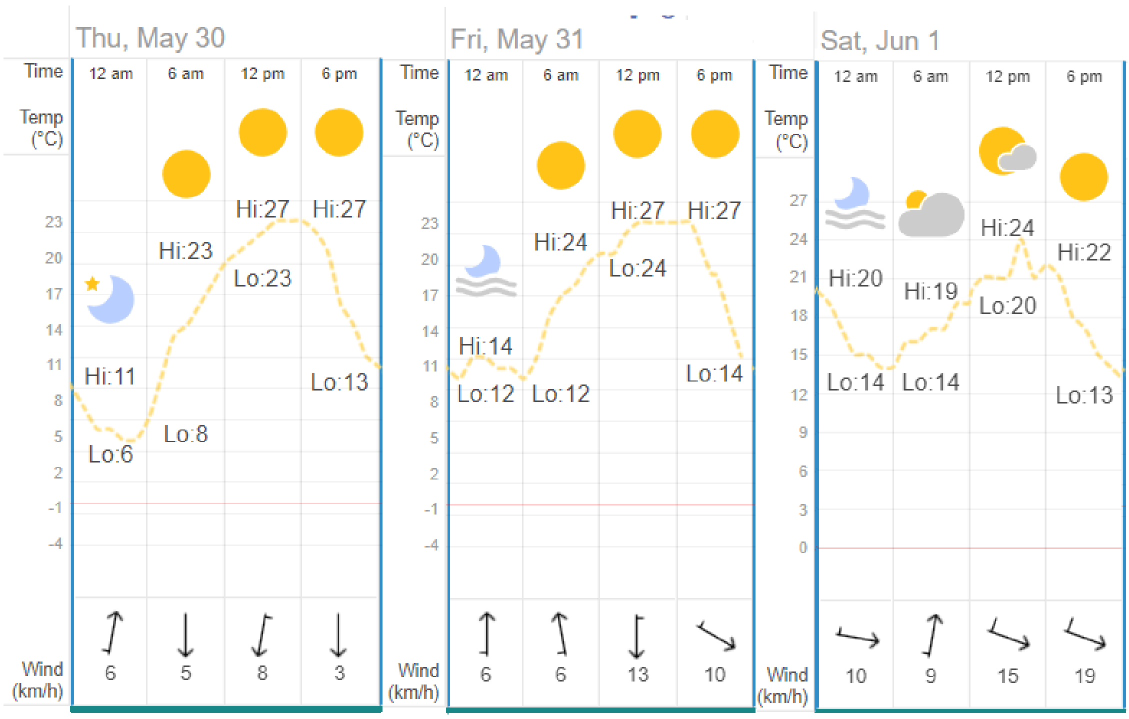


Figure A.6: Sunlight Data in Prince George from 30/05/2019 to 01/06/2019

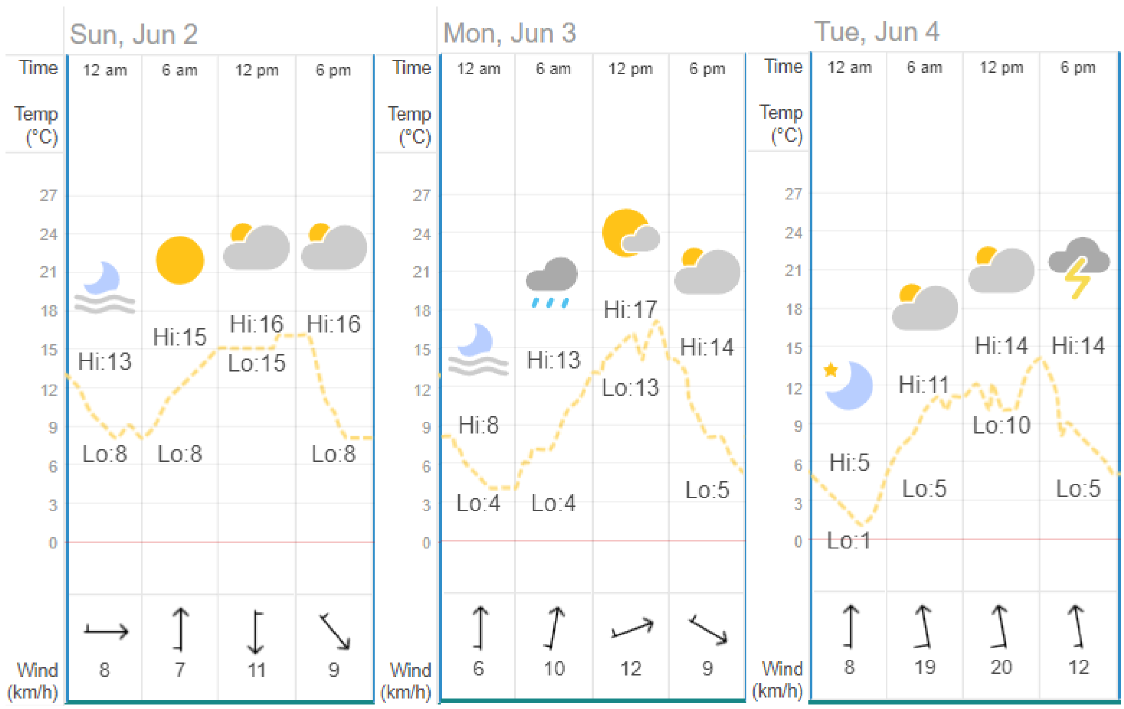


Figure A.7: Sunlight Data in Prince George from 02/06/2019 to 04/06/2019