ASSESSING THE EFFECTIVENESS OF LOW-COST SENSORS AS PROBABLE ALTERNATIVES FOR CONVENTIONAL AIR MONITORING DEVICES

by

Issa Adegbola Tijani
Bachelor of Science, University of Lagos, 2016

Master of Science, University of North Dakota, 2023

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

In partial fulfillment of the requirements

for the degree of

Master of Science in Environmental Engineering

Grand Forks, North Dakota
August
2023

APPROVAL PAGE

This thesis is submitted by Issa Tijani in partial fulfillment of the requirements for the Degree of Master of Science (Environmental Engineering) from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

____

Dr. Yeo Howe Lim

____

Dr. Van der Watt Johannes

____

Dr. Taufique Mahmood

This thesis is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

____

Chris Nelson

Dean of the School of Graduate Studies

____

Date

PERMISSION

Title: Assessing the Effectiveness of Low-cost Sensors as Probable Alternatives for Conventional Air Monitoring Devices.

Department Environmental Engineering

Degree Master of Science

In presenting this thesis, in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Issa Tijani

07/05/2023

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to everyone who has supported me throughout my program at the University of North Dakota. Without your support, guidance, and encouragement, completing this program may not be possible.

I would like to express my profound gratitude to Dr. Yeo Howe Lim for giving me the opportunity to demonstrate my capabilities. My conversation with you before the start of my master's program in Environmental Engineering has been the catalyst that keeps propelling me for academic excellence. Your continuous support, expertise, and leadership have been helpful in shaping my thesis direction and enhancing my research skills.

I would like to acknowledge both Dr. Taufique Mahmood and Dr. Vander Watt Johannes, the knowledge that I gained in your classes has been the backbone of this work. Your guidance has been invaluable in conducting data analysis and understanding air pollution problems.

I would like to extend my appreciation to Allegheny County Health Department, Monitoring Section for giving me access to the necessary resources to establish this thesis. The resources you provided have contributed to the quality of this work.

Lastly, I would like to acknowledge the contributions of numerous research articles, books, and scholarly resources that have formed the foundation of this thesis. Your contributions have guided the exploration of this chosen topic.

DEDICATION

This research work is dedicated to God for being the source of my need, the most beneficent and the merciful. Also, to my family and friends, for their support and care throughout my program at the University of North Dakota.

TABLE OF CONTENTS

CHAPTER THREE

CHAPTER FOUR

LIST OF FIGURES

LIST OF TABLES

ABSTRACT:

In November 2022, Allegheny County residents raised concerns about increased air pollution, especially in the Clairton region area. Due to the recent air inversion, the Clairton region was regarded as a highly polluted area because of the concentrated hydrogen sulfide smell invading the area and the exceedance of the daily PM2.5 average. Consequently, Allegheny County Health Department installed low-cost sensors (PurpleAir) in Mon Valley regions to increase the air monitoring rate and verify the efficiency of federal Reference Methods in place. Low-cost sensors are a relatively new technology adopted in air monitoring. Therefore, this project seeks to assess the effectiveness of these sensors relative to the reference method (FRM) while considering only PM2.5 data for the two methods. According to the EPA, establishing a relationship between the sensor and FRM data requires plotting the variables on a correlation graph and generating an equation of a line with a coefficient of determination (R^2). The coefficient of determination is given a value of 1 and the closer the value of R^2 generated from the plot is to 1, the more likely the sensor behaves like FRM. According to the analysis of the data collected from Allegheny County Health Department using Microsoft Excel and Minitab, R^2 = 0.52, 0.69 and 0.12 for three different sensors. Then, a correction was made to remove the effect of relative humidity consequently, R^2 = 0.63, 0.67 and 0.09. Also, during the analysis, it was observed that the sensors did not generate data for a certain number of days, and as a result, the corresponding number of days were removed from the FRM data to enable correlation. For better visualization, an air quality index (AQI) was generated for the sensors data with corresponding AQI categories, and a heat map was developed for the sensor within the study area. Finally, analysis shows that the first-second sensors established moderate relationships with the reference instrument and the third sensor did not establish any relationship. The sensor was able to detect the presence of pollution but could not generate accurate concentration. This work could not establish why sensor three has a low R^2, but an assumption was made to conclude that the poor performance of the third sensor is either due to the fact that the sensor is bad or early degradation occurred and probably the effect of an unforeseen environmental factor. Therefore, more work needs to be done in the future so that the sensors can be reassessed during other seasons to compare their performance.

Key words: Air inversion, Average, PurpleAir Sensor, Federal Reference Method, Criteria pollutants, EPA, Data collection, Data Analysis, Coefficient of Determination (R^2), T640, Allegheny Health Department**.**

Abstract word count: 415

CHAPTER ONE

1.1 BACKGROUND OF STUDY

Due to the series of environmental issues that have happened in the past, the federal government established environmental laws of which The Clean Air Act is among the important ones. The Clean Air Act was established to ensure that we attain a cleaner atmospheric environment and greater air quality. The Environmental Protection Agency (EPA) was authorized to administer standards for pollutants that have been deemed dangerous to human health. These pollutants are called criteria pollutants. To ensure that we are in attainment with the standards established for criteria pollutants, constant air monitoring, and quality assurance are often required. The Environmental Protection Agency established the Federal Reference Method (FRM) that must be adopted in air monitoring and data collection. These methods required a lot of technicality and series of maintenance and in fact, everything about it has a procedure that must be followed, and as a result, the accuracy of FRM was attained. Recently, a new technology called low-cost sensor was introduced into the air monitoring process. The cost of purchasing them is relatively cheap with fewer technical requirements for installation and data can be generated in real time. Therefore, this project seeks to assess the effectiveness of low-cost sensors relative to the federal reference method.

The Federal Reference Method (FRM) is the main conventional sampling techniques established for collecting and testing PM 2.5 concentrations. FRM combines gravimetric method designed to measure particulate matter concentrations over a period of 24 hours. The gravimetric term means that it’s considered the total mass in a sample to evaluate PM 2.5 concentration. This method involves the use of sampling head equipped with an inlet to draw in air from the atmosphere and remove debris and large particles. The second layer after the sampling head is the cyclone with the capacity of only retaining particles with diameter of 2.5 and rejecting particles greater than 2.5 microns (EPA, 2022).

Historically wise, the Pittsburgh area was known to struggle for cleaner air. In 1941, the City of Pittsburgh passed a law to reduce coal production, but the law didn't get enacted completely until after World War II. While this was a good idea, other factors such as introducing natural gas, switching from coal to diesel for railways and stoppage of steel and iron production in the region quickly improve the air quality of Allegheny County (Byrnes, 2012).

Towards the end of the year 2022, there was a lot of complaint in the news about increased daily PM 2.5 and people with sensitive health issues were advised to stay indoors as a result Allegheny County Health Department deployed Purple Air monitoring sensors in Mount Valley region to support the current monitoring system in place. According to the Health Department, they stated that the increase in air pollutants was due to an air inversion. Air inversion was a switch in the atmospheric process between the lower and higher altitude thereby preventing gas exchange within the atmosphere. More so, low-cost sensors are relatively cheap, and affordable, and generate readings of any range with little and no technical skill required for installation. It was reported that low-cost sensors can only measure the mass of particulate matter by inferring from light scattered by particles and thus, their measurement can be affected by particle sizes and refractive index. More so, the relative humidity would have much on its particulate matter mass measurement and recommended to always correct for relative humidity in case of exceedingly high measurement. Hence, this project seeks to evaluate the effectiveness of the low-cost sensor as a probable alternative for air quality monitoring devices.

1.2 STATEMENT PROBLEM

In November 2022, it was broadcasted that the Pittsburgh area exceeds the daily PM2.5 average and people with challenging health conditions may be in danger consequently, Health Department, deployed low-cost sensors in the Mon Valley region to augment the initial monitoring system in place. Therefore, this work seeks to analyze the effectiveness of the sensor data relative to the reference method in place.

1.3 OBJECTIVES

1. To review an existing literature on the possibility of using low-cost sensor as an alternative to the federal reference method
2. To find out if truly PM2.5 daily average was exceeded and not just a rumor.
3. To determine if there is a correlation between Sensor data and Federal Reference Method data.
4. To identify which air quality parameters can be effectively measured by low-cost sensors.
5. To determine if low-cost sensors could be recommended for regulatory purposes or individual uses.

1.4 RESEARCH QUESTIONS

1. How comparable is the accuracy of low-cost sensors measurement to the conventional air monitoring devices?
2. What are the limitations of low-cost sensors compared to the federal reference method?

3. How does low-cost sensor compare in terms of cost and maintenance to the federal reference method?
4. Can we improve accuracy of these sensors by calibration?
5. What are the benefits of using these in different environmental settings?

CHAPTER TWO

1.0 INTRODUCTION

2.0 LITERATURE REVIEW

With growing concerns for exposure to PM 2.5, with little availability of spatiotemporal data to evaluate exposure to particulate matter, and with the cost of conventional monitoring systems, makes it invaluable to explore low-cost as an alternative to supplement the existing monitoring systems. Preexisting research shows that when optical monitors were collocated with Shinyei PPD42NS in the United States, a strong correlation was observed between them but when the sensor was collocated with a gravimetric reference method, moderate correlation was observed. Even though one of the correlations was moderate still, the two observations were able to identify pollution hotspots. Many low-cost sensors including PurpleAir have been deployed worldwide to locate pollution hotspots more spatially than they do for monitors used for regulatory services (Gao *et a*l., 2015).

Four different low-cost sensors were calibrated with a TEOM reference method under steady conditions and observed that when the particle sizes were kept constant in the laboratory, the sensors generate a perfect correlation with the reference method. The low-cost sensors did not measure the mass of PM 2.5 directly rather, they measure particle size distributions under the assumption that particle sizes are spherical, and density is known therefore, error in measurement can easily be magnified. The sensors employed were based on the photometric method, but the operation conditions were steady instead of transient as known in other studies. The Lag time between sensor responses and reference can result in a potential error during calibration therefore, selecting the right reference method for calibration is crucial. (Di Liu *et al.*, 2017).

Different studies have documented spikes in low-cost sensor measurements and therefore, there is a need for a way to calibrate the monitors for better efficiency (Malings *et al.,* 2019). Kelly *et al*. (2017) observed that PurpleAir model PMS1003 reported elevated PM 2.5 concentration when atmospheric particulate matter concentration exceeds 10 μg/m^3 in the United States during winter air inversion. Since the low-cost sensor principle of measuring PM is based on a light scattering mechanism hence, calibrating it will be difficult because the detection of PM concentration may be affected by changes in particle sizes and therefore, calibration developed in one place may not be applicable in another place (Sardar *et a*l., 2005).

Malings *et al*. (2019) explained ways to correct low-sensor data. The first correction was used to correct for the major factor contributing to the major difference in measurement between PurpleAir and T640 by removing the effect of relative humidity, a hygroscopic growth factor. The second correction was to correlate the size distribution, chemical composition of factory calibration and the aerosol to be measured. Even after the correction, a large difference was still observed between PurpleAir and reference data hourly average. When an aerosol composition is close to this average value, the empirically derived equation would work well but if the composition is not close, it won't work well.

Jessica *et al*. (2020) reported that particulate matter concentrations measurement taken with low-cost sensors were often elevated. They investigated this by attempting to calibrate the PurpleAir sensor using field and laboratory approaches by collocating PurpleAir sensors with a TEOM reference method. They established linear relationships between the two sets of data and

R^2 = 0.99 but F-test indicated there was no fit between the model and the data. They observed that the correction factor developed in the laboratory does not translate to the field, but the correction factor derived from ambient relative humidity improved PurpleAir PM 2.5 concentrations that are within 20% of the TEOM PM concentration from 24% to 66%. Data from sensor networks can be very useful to people that are allergic to particulate matter exposure (Nelson, 2016).

(Wallace *et al*. 2021) compared data from 33 outdoor PurpleAir sensors to a nearby 27 EPA Air Quality Stations over a period of a year and a half. The sensors with ALT_CF3 algorithms had better precision and lower detection limit. A calibration factor was developed for the sensors by comparing all its data to the reference data and a very strong correlation was achieved. Sensors high reading value correspond to a major fire event, each event had different correlations factors and the observation shows that sensor data were overestimated by 40 % which agrees with other researchers about overestimation of particulate matter data reported by low-cost sensors.

(Wallace, 2022) compared functionality of the recent low-cost sensor (ALT_CF3) algorithm to the earlier version algorithm (CFI_ATM) in an attempt to answer questions regarding the reliability of the sensors in acquiring data, sensor degradations, and accuracy of the algorithms in estimating PM 2.5 concentrations. He observed that CFI_ATM lost about 25 – 50 % of data in an attempt to convert data below a certain threshold to zero and no data was lost using ALT_CF3 with 92 % better precision than the CFI_ATM. He concluded that both algorithms are not good for indoor monitoring and low atmospheric concentrations but the ALT_CF3 made by PurpleAir can be used.

2.1 INTRODUCTION TO AIR SENSOR

The term low-cost sensor originated from advancements in microprocessing technology, and they have been referred to as Original Equipment manufacturer Devices (OEM) or Sensor Systems or Sensor Nodes in literature. Air Sensors are a piece of equipment capable of sampling ambient conditions and generating readings of high frequency such as in seconds and minutes. They are capable of measuring most air pollutants but not all. They generate quick readings with high variation in data quality hence, differences may exist in measurements taken by sensors of the same model (EPA, 2022). The figure shows typical Air Sensor components.

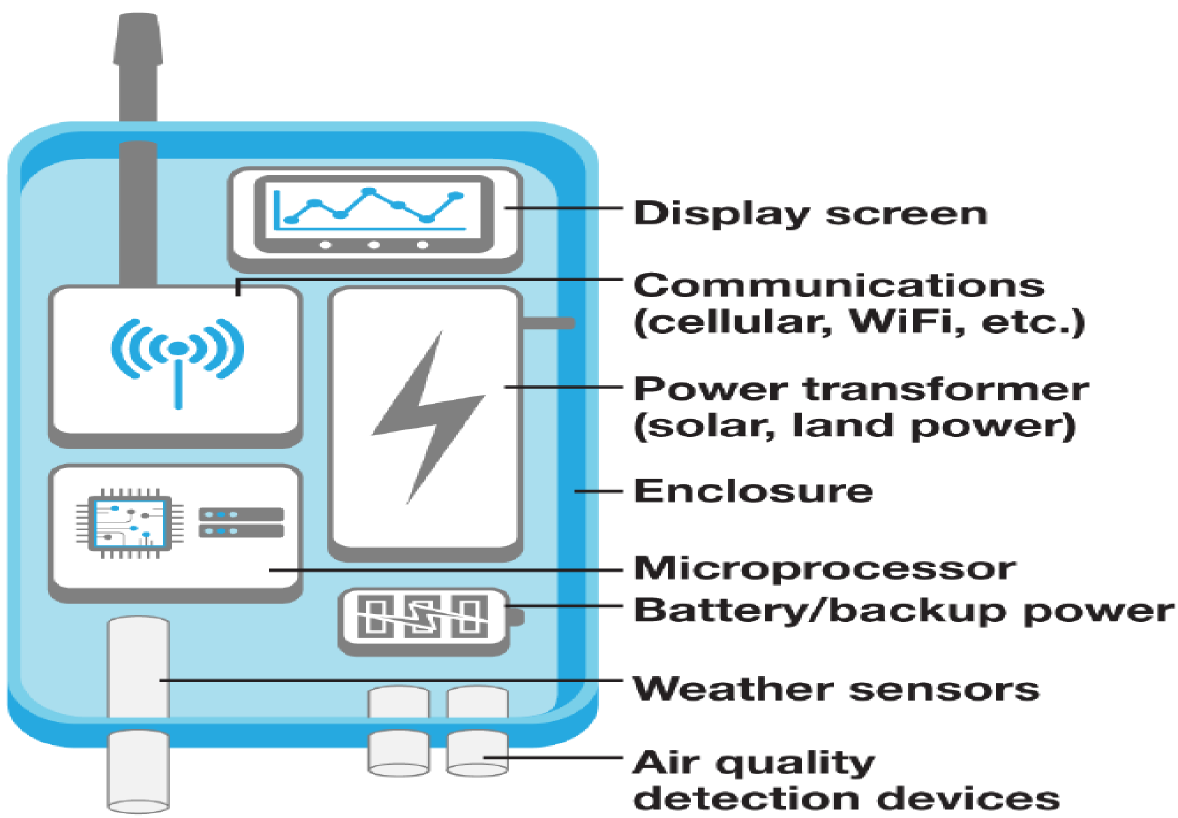


Figure 1.1: Typical Air Sensor (EPA, 2022)

The Environmental Protection Agency has established that air sensors cannot be used for regulatory services but may be used for information monitoring purposes. Since sensors do not meet conventional monitoring criteria, it may be essential during spatiotemporal analysis either to monitor air quality trends of a geographical location or to use as a tool for air forecasting. It was also established that air sensors may be invaluable to locate pollution hotspots especially when certain regions increased air pollution more than the others and it may be applicable to epidemiological studies to visualize changes in air pollution over time (EPA, 2022).

2.2 AMBIENT ENVIRONMENT

The word air quality is used to illustrate how much pollution is in our atmospheric environment. Air quality has become a subject of concern because it has been attributed to many health problems such as asthma and respiratory diseases as demonstrated by epidemiological reports. Furthermore, it can reduce visibility, destroy organisms' lives, and to the point of damaging climate conditions. Air pollution comprises solid particles, liquid droplets, and gases. It can be created by human activities called anthropogenic sources such as emissions from electrical power plants, vehicles, and oil rigs. Natural sources include wildfires and volcanoes. Air quality is not stable and changes depending on weather conditions. The lifespan of any air pollutant depends on its reactivity with other factors such as wind, temperature, relative humidity, sunlight, and precipitation in the atmosphere. Excessive wind speed can decrease air pollutants concentration by dispersion over time. Another type of pollutant that has caused a huge spike in data recorded at monitoring sites in Allegheny County is the stagnant pollutant called air inversion. Air inversion is a phenomenon whereby cool air is trapped near the surface and warm air stays above thus, no exchange can occur. Consequently, pollution builds up near the surface and could remain like that for days. See Figure 2.2 below for details. (EPA, 2022)

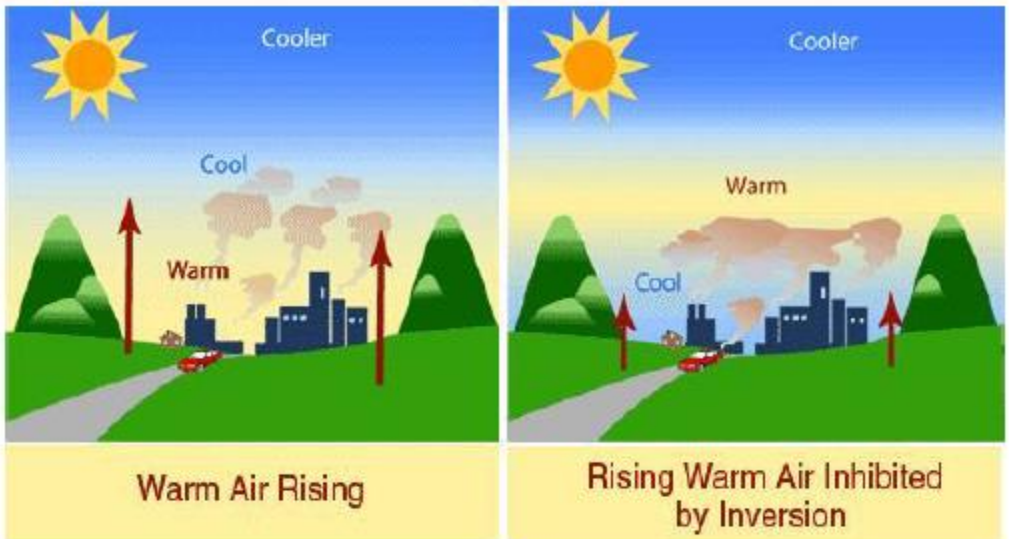


Figure 2.2. Typical airflow versus trapped air (Air inversion): (EPA, 2022)

Pollutants can either be primary or secondary. The primary pollutants come directly from a source while the secondary pollutants are formed after reacting with the atmosphere and their concentration varies depending on weather conditions, emission rate, and closeness to a source. The concentrations are either reported in ($\mu g/m^3$), ppm, or ppb. Table 2-1 below represents criteria pollutants.

Table 2-1. Criteria pollutants (EPA)

Pollutant [links to historical tables of NAAQS reviews]		Primary/ Secondary	Averaging Time	Level	Form
Carbon Monoxide (CO)		primary	8 hours	9 ppm	Not to be exceeded more than once per year
			1 hour	35 ppm	
Lead (Pb)		primary and secondary	Rolling 3 month average	0.15 μg/m^3 (1)	Not to be exceeded
Nitrogen Dioxide (NO_2)		primary	1 hour	100 ppb	98th percentile of 1-hour daily maximum concentrations, averaged over 3 years
		primary and secondary	1 year	53 ppb (2)	Annual Mean
Ozone (O_3)		primary and secondary	8 hours	0.070 ppm (3)	Annual fourth-highest daily maximum 8-hour concentration, averaged over 3 years
Particle Pollution (PM)	$PM_{2.5}$	primary	1 year	12.0 μg/m^3	annual mean, averaged over 3 years
		secondary	1 year	15.0 μg/m^3	annual mean, averaged over 3 years
		primary and secondary	24 hours	35 μg/m^3	98th percentile, averaged over 3 years
	PM_{10}	primary and secondary	24 hours	150 μg/m^3	Not to be exceeded more than once per year on average over 3 years
Sulfur Dioxide (SO_2)		primary	1 hour	75 ppb (4)	99th percentile of 1-hour daily maximum concentrations, averaged over 3 years
		secondary	3 hours	0.5 ppm	Not to be exceeded more than once per year

2.3 OUTDOOR AIR MONITORING

Air Pollution Monitoring involves identifying different pollutant concentrations in the atmosphere by determining the magnitude and classes of a particular pollutant detected during observation. Different methods and equipment are deployed to quantify the pollutants and therefore, it is paramount that we select an instrument that can provide accurate, consistent, and traceable measurements. The different measurement categories are discussed below:

REFERENCE MONITORS: They are used to assess compliance with National Ambient Air Quality Standards (NAAQS). It can either be a Federal Reference Method (FRM) or Federal Equivalent Method (FEM) and must adhere to requirements stated in the U.S. Code of Federal Regulations (40 CFR Parts 50, 53, and 58). The reference method generates highly reliable and accurate data. The FRM particulate matter used filters to collect particles over a period whereas, the FEM continuously monitors particles say every hour, and the list of recognized FRM/FEM monitors is presented in U.S EPA Ambient Monitoring Technology Information (AMTIC). The accuracy of FRM relies on constant precision checks and quality assurance and control.

RESEARCH INSTRUMENTS: These are different types of air monitoring devices developed for a special purpose to achieve certain results in the field. They are often created because there is no existing method to be used. They are often operated by experts and can be incorporated into FRM/FEM if they show comparable capability in measurements.

REMOTE SENSING: This method does not require physical touch and measuring can be taken from a distance. It involves measuring emitted light in order to collect information about a specific pollutant. It is a valuable method to detect particulate matter, gaseous pollutants, and certain volatile organic compounds. Data can be collected using aircraft, satellites, or ground-based installations.

AIR SENSORS: See section 2.1.

Table 2-2. Differences Between Reference Monitors and Low-cost Sensors (EPA, 2022)

Parameters	**Reference Monitors**	**Low-cost Sensors**
Equipment cost	Between ($15,000-$40,000)	Between ($100-$5,000)
Personnel	Required highly trained personnel	Understanding the user guide is sufficient
Cost of Maintenance	Very expensive (cost of housing the equipment, trained personnel, repairs, and quality control)	Inexpensive to replace the sensors or repair and maintained
Sitting Location	Instruments required a permanent location such as providing a shelter	It is very portable and therefore, it can be transported
Lifespan	≥ 10 years	Between 1 to 2 years
Regulatory Monitoring	Used for regulatory services	Supplementary purposes

The location of the air monitoring network is very important because the closer the air quality monitors to the source of the pollution, the higher the concentration of the pollutant recorded and vice-versa. Even if pollutants are distributed over a given area, the concentration of the pollutant recorded at the monitoring stations can be differentiated by location. Therefore, geographical locations and ambient conditions could influence pollutant concentrations. An air pollution monitoring network comprises various sites with instruments to measure one or more pollutants. These networks are purposefully designed and operated to fulfill specific objectives such as using FRM or FEM instruments to achieve the requirements stated in the National

Ambient Air Quality Standards (NAAQS) as stipulated by Clean Air Act (CAA). Accuracy and completeness are very important to the attainment of National Ambient Air Quality Standards. "The U.S. Environmental Protection Agency AirData Air Monitoring website provides an interactive visualization of air monitoring locations and provides monitors-specific information for different regulatory air monitoring networks using the AirData Map." (U.S. EPA, 2022).

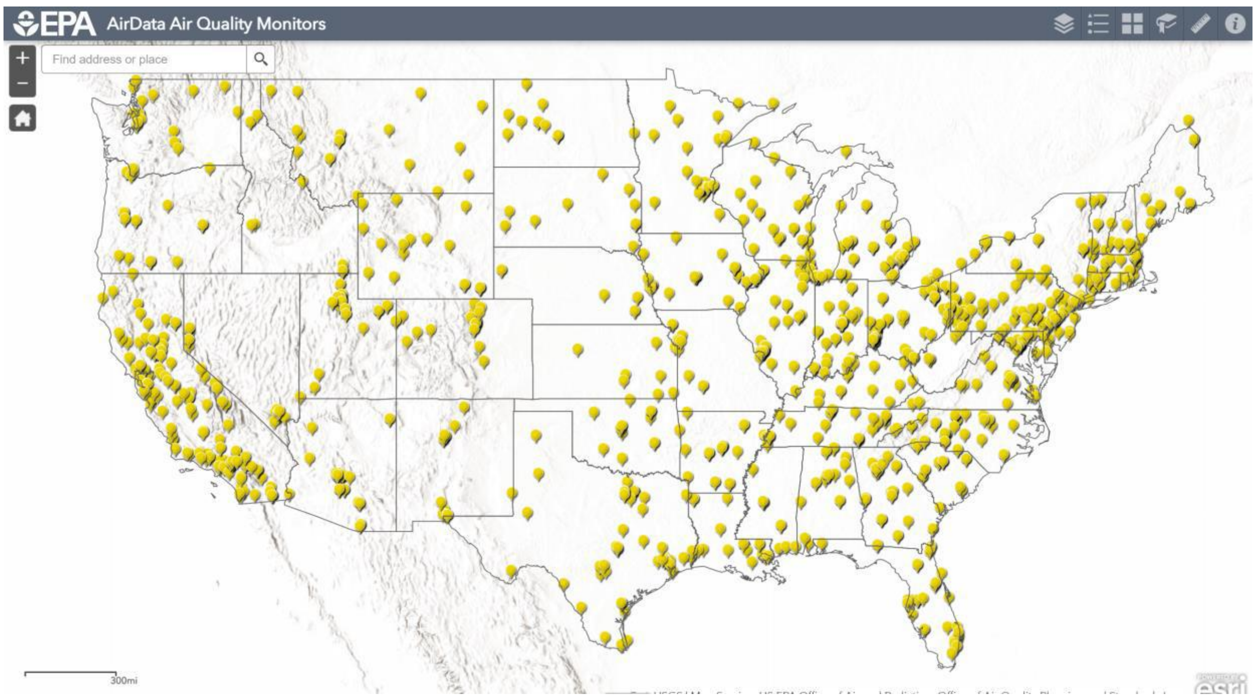


Figure 2.3. U.S. EPA AirData Air Quality Monitors Website for PM 2.5 Continuous Monitoring Stations (U.S. EPA, 2022).

Nationwide, there are over 4000 stations that are monitoring criteria pollutants such as Particulate Matter, Ozone, lead, Carbon monoxide, Sulfur dioxide, and Nitrogen dioxide using either Federal Reference Method (FRM) or Federal Equivalent Methods (FEM). To ensure accurate measurements, different agencies in charge adhere to strict standards such as operational plans, and quality assurance including precision checks, calibrations, audits, performance evaluations, and data validations (U.S. EPA, 2022). All air monitoring agencies submit all their quality- assured data to EPA's Air Quality System (AQS) database.

It is very difficult to quantify the impact of air quality on health, despite this difficulty, government agencies established different ways to put this impact on text. For an instant, The U.S.EPA did Integrated Science Assessments, the 2021 Air Quality Report prepared by European Union, The World Health Organization also established a standard but here in the United States, we abide by the following standards: National Ambient Air Quality Standards (NAAQS), Air Quality Index (AQI) and National Institute for Occupational Safety and Health (NIOSH).

National Ambient Air Quality Standards (NAAQS): In 1990, when the Clean Air Act (CAA) was last amended, Environmental Protection Agency was required to establish standards for criteria pollutants (Carbon Monoxide, Lead, Ozone, Sulfur (iv) oxide, Particulate matter (PM 2.5 & 10). These pollutants have been considered for human health and the CAA identified primary and secondary pollutants. The standards are to protect our health including the sensitive groups and the secondary standard is to protect our resources (See table 2.1 for details).

Air Quality Index: This is a numerical scale established by the EPA to translate air quality of any location to the people. The scale was divided into six categories and each category signifies different health conditions. EPA established this scale for all criteria pollutants except for lead. The color code is to notify the public of the health condition of their vicinities: See table below.

Table 2-3. The Air Quality Index

AQI Color	AQI Category	Index Value	Air Quality Status
Green	Good	0 - 50	Good condition with little or no risk associated.
Yellow	Moderate	51 -100	The air quality condition is acceptable but may be risky to hypersensitive people.
Orange	Unhealthy for Sensitive Groups	101 -150	The air quality is unsafe for sensitive group members although, the general public may not be affected.
Red	Unhealthy	151 -200	Some members of the public may be affected: sensitive group members may experience bad health conditions.
Purple	Very Unhealthy	201 -300	Health alert: Bad health conditions may increase for everyone
Maroon	Hazardous	≥ 301	Emergency health warning and everyone is likely to be affected.

With the exception of lead, EPA has developed AQI for most criteria pollutants. The AQI scale ranges from 0 – 500, with the highest value signifying bad health conditions.

Table 2-4: Demonstrates how sensitive Groups Respond to Different Polluted with AQI Greater than 100 (“AirSensor Guide, 2022”)

Pollutants	**At-Risk Populations**
Carbon Monoxide (CO)	People with heart disease
Nitrogen Dioxide (NO2)	People with asthma, children, and older adults
Ozone (O3)	People with lung disease, children and teenagers, older adults, people who are active outdoors (including outdoor workers), people with certain genetic variants, and people with diets limited in certain nutrients.
Particulate Matter (PM2.5)	People with heart or lung disease, older adults, children, and people of low socioeconomic status.
Particulate Matter (PM10)	People with heart or lung disease, older adults, children, and people of low socioeconomic status
Sulfur Dioxide (SO2)	People with asthma, children, and older adults

Exposure to different pollutants causes different reactions in our body therefore, EPA calculates AQI values for different pollutants on the average of 1, 8, or 24 hours. For instance, exposure to ozone can cause respiratory problems within an hour to a day and exposure to SO2 can cause the same problem within 5 minutes (EPA, 2022).

The FRM and FEM instruments are the piece of equipment used to check for compliance with National Ambient Air Quality Standards (NAAQS) therefore, it is important to ensure that measurements are aligned to this standard when comparing air measurements and, ensure that concentrations are corresponding to specific time range. For instance, to compare the low-cost sensor measurements to that of NAAQ. Ensure that the sensor data is cleaned and corrected and matched with the corresponding NAAQS data averaging time (EPA, 2022).

National Institute for Occupational Safety and Health (NIOSH): This body creates policies and suggestions for work-related injury and illness, including exposure to air pollutants. The guidelines typically focus on shorter time frames as they are specific to work-related environments and schedules such as workdays (EPA,2022).

2.3 Low-cost Sensor Monitoring

Using a low-cost sensor for air monitoring can be very difficult and achieving good results requires good planning. This planning will enable good data collection, trusted data and make available data suitable for any desired purpose and most importantly save time and money.

It is very important to document the questions you wish to be answered before starting any air monitoring processes. A simple example could be to investigate the PM 2.5 concentration of a vicinity and corresponding changes between day and night. Having clearly stated questions provides a roadmap to establish good data collection, the type of sensor to be used, data processing, data validation, and quality assurance. An area where the network of sensors will be installed should be free of interference and possibly should be collocated side-by-side with a reference instrument. Before data collection, ensure that site visitation is done to ensure that the sensors are working perfectly. Finally, the approach adopted in analyzing or interpreting the result is paramount to achieving the air monitoring goals. Certain cases do not require planning such as when we are not interested in knowing the absolute concentration of a pollutant especially when we are interested in knowing when the air quality is good or bad (EPA,2022).

2.31 Questions for Determining the Purpose of Monitoring

Finding out which dataset is already available and why we need new datasets is important in determining which sensor to purchase. Asking necessary questions would help to identify pollution of interest, condition of the field, period of data collection, and quality of measurement desired. Another question to keep in mind before proceeding with air monitoring is what do you intend to do with the data that will be collected? Are you planning to share your findings with certain stakeholders? Is there any information you hope to share with others? Are there any potential actions that need to be taken? These types of questions will help you determine the type of team needed for the project and the type of data quality to be acquired. Do not wait for data to be collected before deciding its purpose (EPA,2022).

2.32 Developing a Plan

Developing a project plan is crucial to preventing complications in the future project. It is important to emphasize the pollutants of concern and discuss likely environmental factors that may affect our results. The likely environmental factors to be monitored such as temperature, relative humidity, wind speed, and wind direction should be discussed. The location to install the monitors and how the data collection will be assessed should all be examined. A detailed project plan will indicate completeness in the process and will ensure that project objectives are met. A plan should help pinpoint any potential problems ahead of time therefore, the plan should be shared with any available experts for constructive feedback. The plan should also be shared with those that the results are meant for. A properly designed plan would inform the stakeholders of the decision before investing their time and money. Thus, A Quality Assurance Project Plan is a written document that shows how an organization will adopt quality assurance and control activities such that data collection will be of high quality and can be used for intended purposes. The table below describes the topics that must be covered when excluding a project plan.

Table 2-5: Step by step methods of establishing project plan (EPA 206-B-18-001, 2019)

Topics	Purpose
Purpose of Monitoring	State the problems that need to be addressed and the expected outcome to be achieved using low-cost sensors
Project/Task Organization	Identify the role of every personnel that will be involved in the project.
Engagement with Local Partner	Collaborate with other experts such as local agencies and discuss your project with them and your expected result.
Project/Task Description	Work summary, objectives, work schedule, and expected outcome.
Data Quality Objectives	a) Why do we need the data?

	b) Are they already available? c) What measurements are required and what should they signify? d) Is specific accuracy required?
Contingency Planning	Create an alternative plan in case something goes wrong. For instance, a) If staff resign b) Dealing with sensor failure c) The vandalization of a site and equipment d) Develop many possible outcomes and plan for troubleshooting
Training and Experience	Identify any potential training needed or certification to complete the project.
Documentation and Records	Determine how field activities will be recorded either on the logbook or SOP forms.

2.4 Questions to Ask When Selecting Low-cost Sensor

Before selecting a low-cost sensor for air monitoring, certain questions need to be asked. Answering these questions will enable us to choose the right sensor that will meet our project objectives and will enhance good data collection (EPA,2022).

1) What do we need the sensor for?

a. Is this sensor for educational purposes?

b. Is it for hotspot identification?

C Is it to determine personal exposure or participatory science?

2) Which pollutants are we trying to detect?

a. Is it a gaseous pollutant such as ozone and nitrogen dioxide?

b. Is it particulate matter (PM 2.5 & PM 10)?

3) **Which features are we looking for in the sensor**?

a. Is it resistant to various weather conditions?

b. Is it the dimension, weight & portability?

c. Is it an energy supply?

d. Is it memory capacity?

e. Is it maintenance requirements?

4) How do we assess the performance of this low-cost sensor?

a. Correlate the result to a regulatory station.

b. Carry out periodic quality assurance checks.

c. Check for meteorological and other factors that may impact performance.

d. Regularly review and analyze data to pinpoint any potential errors.

5) How much does a typical air sensor cost?

Capable of measuring 1 -2 pollutants may range ($ 150 - $1,500)

1 -3 pollutants may range ($500 - $2,500)

4 pollutants & above may range ($2,500 - $10,000)

6) **Which information should I look for in a manual?**

Which pollutants can the sensor measure?

What are the general operating instructions?

How do you assess the information gathered with the sensor?

What are the expected results?

How do I reach out to customer support?

2.5 Working Principle of Low-cost Sensor

Particulate Matter Sensors: These sensors typically use an optical method whereby particle mass concentration is determined by the amount of light scattered by the particle. The scattering amount is greatly influenced by particle size, shape, and chemical composition. However, this method can only detect a very small range of particles size therefore, particles can build up over time and thus, leading to poor sensor performance. The particulate matter sensors can last between a year to 4years and can be used for outdoor, indoor, and smoke monitoring purposes. Many of these sensors cannot detect PM10 and dust.

Gas Sensors: This sensor is commonly used for Ozone, Nitrogen dioxide, and Sulfur dioxide pollutants detection. It works on the principle of using electrochemical or metal oxide sensors for detection. It can respond to meteorological factors such as temperature, relative humidity, and other interference gases. Regardless of usage, this sensor tends to experience a reduction in sensitivity over time and thus, necessitating regular collocation at least seasonally. The response of this sensor in low concentrations often poses challenges and the most successful application was reported near to the source. The lifespan is usually between 6 months to 2 years.

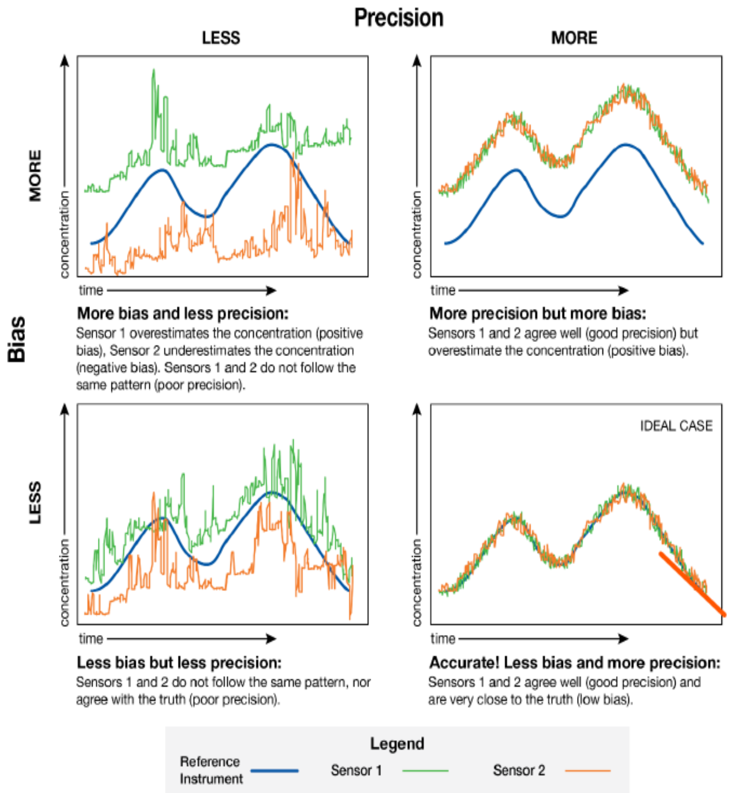
Total VOC (tVOC) sensors: This sensor can either be a metal oxide semiconductor (MOS) or a photoionization detector sensor (PID). These sensors have the capability to measure many volatile organic compounds (VOCs) and they may respond to some compound than other as a result, the VOCs concentration reported by this sensor does not represent through value and cannot be attributed to a particular compound. They also respond to temperature, relative humidity, and other gases that may interfere with measurements. The most successful application was near the source when comparing the upwind to downwind to detect a spike in VOC emissions. The PID can last for months up to a year while MOS have a longer lifespan.

2.6 Measurement Range and Detection Limit:

The range between the lowest and the highest measurement taken by a sensor is referred to as Measurement Range. A sensor becomes most helpful if it can measure the complete range of the observed pollutant found in the atmosphere. Depending on the closeness to the observed source, the ability of a sensor to detect very high or very low pollutant concentrations is very important. The detection limit is the lowest observed concentration that can be measured by an instrument on a regular basis. It is crucial to refer to manufacturer specifications to ensure that the sensor can measure the detection range found in the U.S. (EPA,2022).

Sensor Accuracy*:* It is used to describe the relationships between sensor measurements and measurements from a reference monitoring station. Bias and precision are the two important components used to explain accuracy.

PRECISION*:* This described how well a sensor can generate the same result as produced by reference measuring instruments.



BIAS: It represents error in sensor measurement. See figure 2.6 below for details.

Figure 2.6: Descriptions of Accuracy, Precision and Biasness in Low-cost Sensor (EPA, 2022)

Calibration is a process used to adjust instrument settings such that measurement can be matched with certified standards for accurate comparison.

Collocation is a way of assessing the performance of low-cost sensors by installing them near a reference method instrument.

Data correction: Involves adjusting low-cost sensor data to improve its accuracy relative to reference data.

CHAPTER THREE

3.0 METHODOLOGY

3.1 SITE SELECTION

Selecting a site to install low-sensor either a single one or a network of low-cost sensors is very important. The process of finding a good site will ensure that the sensors are collecting a representative sample of the surroundings. Aside from the placement of the sensors, factors such as availability of electric power, internet access, security of the site and access to carry out necessary maintenance are crucial.

The process involved in identifying the right location for air sensor installation requires a complete analysis of the environmental conditions. This includes assessing the topography, prevailing atmospheric conditions, wind directions, pollution sources, and any other factors that may contribute to air quality problems. By strategically selecting a good site, the data collected will be a representative sample that can be used for many purposes. Hence, site selection is beyond just sensor installation but rather an attempt to ensure that the data collection process is accurate and generate a representative sample (U.S. EPA, 2017).

One of the most important aspects of the site selection is the availability of electric power to ensure that the sensors are collecting data continuously without any interruption. Also, where internet access is needed, it is imperative that the site can provide Wi-Fi connections so that real-time data transmission or remote access is possible. Having all these site qualities taken care of will ensure that we are obtaining reliable data and guaranteed an efficient air monitoring system.

Adequate security is required in order to protect the integrity of the monitoring network established with the sensor network. The site location must have adequate security to prevent theft or vandalism and thus, ensure that data collection is protected and can be used with no worries. Easy access to the site is crucial to minimize downtime so that maintenance crews can access any issues relating to the sensors and general operating equipment, hence maximizing the overall efficiency of the monitoring station (U.S. EPA, 2017).

Before selecting a particular location to establish an air monitoring network, certain questions need to be answered in accordance with the monitoring objectives. Since this project is to inspect for changes in particulate matter over the Pittsburgh area, Pennsylvania, United States. The questions listed below were used by Allegheny County Health Department to determine the right location for the sensors' air monitoring network.

Where is the most densely populated region?

Where do we have the most susceptible and sensitive people such as schools and hospitals?

Is there a location with no monitors within the area covered by regulatory monitors?

Is there an area that is not impacted by increased particulate matter concentration?

Where is the area with the maximum particulate matter concentration so that we install air sensors?

Where in the upwind or downwind do we install air sensors to capture the transport and dispersion of particulate matter?

How do we collocate the air sensors next to the regulatory monitors to enable calibration and correction activities?

After considering all the questions stated above, Allegheny County Health Department deployed low-cost in the Mount Valley region because the area meets all the criteria necessary to establish an air sensor monitoring network. See figure 3.1 below for details.

Figure 3.1: North Braddock Air Quality Monitoring Station Map

3.2 SENSOR CALIBRATION

All air monitoring instruments need to be checked more often to ensure that they are working properly and that they are generating highly reliable data. Environmental Protection Agency mandates every regulatory agency to calibrate their equipment according to Quality Assurance Guidance Document. In order to achieve this, regulatory agencies, need to challenge their air instrument with a known concentration or with a known standard, and any adjustment made during this process is then used to set their instruments to match the standard used during calibration. To attain high instrument performance, this process needs to be repeated more often.

Low-cost sensors also need to be calibrated but cannot be done like the reference instrument hence, it is collocated side-by-side to the reference instrument to check for comparability in data collection. Since it is not possible to adjust the air sensor, we then adjust the data to match with reference data. This process of adjusting the data to match the reference data is called Data Correction (U.S. EPA, 2022).

3.3 Air Sensor Collocation

Collocation is often done at pre-existing air monitoring locations throughout the United States however, it may require creating relationships with the local regulatory agency because there may be several restrictions for collocations due to lack of space, electrical power, access requirements, and liability issues. Another option is to work with academic partners or contractors to gain access to a reference instrument relating to your project. The most important factor here is to ensure that all quality assurance procedures are being followed. The sensors should be set up within 20 meters horizontally away from the reference instrument and should be a 1-meter vertical distance from the reference instrument. Most importantly, there should be no obstruction to the airflow between the reference instrument and the air sensors to prevent interference with the data collection.

Figure 3.2. Air Sensor Collocated with Reference Instrument (T640) in North Braddock

3.4 Correction of An Air Sensor

The rate at which the reference monitor report data is important to establish a comparison with the sensor's reporting capabilities. It is important to know how the reference instrument report data because reporting interval may vary depending on the standard in place. Is the reference instrument reporting every 5 minutes, hourly, or 24-hours cycle? This is normally common for filtered-based instruments. For proper comparison, it is imperative that the sensor is able to report data at the same rate as the reference instrument. Practically the sensor should have a slightly shorter reporting frequency than the reference monitor if possible. By selecting sensors with shorter reporting cycles, it is possible to capture more detailed information. Say, for instance, the reference instrument is reporting at 5 minutes interval then choose a 1-minute reporting interval for the sensor then, average the entire first- five minutes of data, and then matched it with the reference 5 minutes data. Combining the reporting intervals will facilitate more detailed analysis and accurate comparisons (U.S. EPA, 2019).

The outcome of collocation can be used to correct the sensor data by matching both reference and sensor data together. To correct sensor data, the first step is to compare the reference and sensor data that was obtained during collocation and ensure that the sampling times match each other and average them out. For instance, the sensor data collected for this project was minutes data therefore, all the minutes data within a specific period were averaged out to match the corresponding time of the reference data. Establishing a correction between the monitors will help to remove the presence of known bias and unseen interference from weather and other pollutants and may be done by developing algorithms or by establishing a simple set of equations.

To establish a relationship between the sensor and the reference data, plot the two data on a correlation graph to visualize it. The sensor data should be plotted on the y-axis (PurpleAir) while reference data should be on the x-axis (T640 data). The Environmental Protection Agency recommends using any analysis tool so far that you are proficient enough to achieve the desired results. The most common correction algorithm equation is the use of the Ordinary Least square Linear Regression Equation. Although R programming could also be used for analysis, EPA did not mandate any specific analysis tools therefore, EPA published the use of Macro Analysis Tools. The relationships between the two monitors correspond to the Coefficient of Determination (R^2). The closer the value of R^2 is to 1, the more the sensor behaves more like the reference instruments (U.S. EPA, 2019).

3.5 Correction Algorithm Equation

The most common correction algorithm uses the equation of the line (Linear Regression Equation).

y = mx + b ------------------------------- (1)

y = PurpleAir sensor measurement

m = slope of the regression line

x = T640 reference data

b = intercept of the regression line

The closer the value of m is to 1, the more the sensor behaves like the reference instrument (T640).

The intercept (b) described biasness in measurement.

Sensor = m * reference + b --------------------------(2)

To move the sensor away from the reference monitor, use the equation below for adjustment.

Adjusted sensor concentration = (measured sensor concentration - b) ÷ m ---------------- (3)

Air sensors relied on light scattered by particles to infer particulate matter concentration hence, changes in particle sizes and refractive index are problematic to low-cost sensors (Austin *et al.*, 2015). Relative humidity can be extremely challenging for air sensors because of changes in particle size distribution and refractive index that may be developed from aerosol water uptake when RH reaches 50 % or more (Malings *et al*., 2019). Calibration of a low-cost can as well be problematic because the particle size used to characterize the PM concentration is not stable and it either be spatial or temporal therefore, calibration made in one place may not translate into other places (Sardar *et al*., 2005). So, one important correction approach is to correct relative humidity using the aerosol water uptake equation.

$CP_{A'} = CP_A / (1 + (0.25\ RH^2) / (1 - RH)$ --- (4)

$CP_{A'}$ = Corrected sensor value

CP_A = Uncorrected sensor value

RH = Relative humidity

3.6 DATA ANALYSIS (CORRELATION & REGRESSIONAL ANALYSIS)

The PurpleAir sensor data that was collected for this project was refined into a useful format using PurpleAir Tool and then exported into Excel and Minitab for analysis. The sensor data were in minute averages therefore, for easy visualization and reducing data sizes, all the data were converted into hourly data. The data were filtered to remove zeros and outliers. On the other hand, the reference method data does not require any processing other than correlating it with the sensors data because the instrument is equipped with the capability of flagging unusual readings for validations. The sensor data was reviewed thoroughly prior to any processing, and it was observed that the sensor did not generate data for some days therefore, corresponding missing data were removed from reference data for easy correlation. According to EPA, a minimum of three sensor data will be required for correlation with reference data therefore, three sensor data were obtained and were used to establishe relationships with one reference data (T640) and all the observations will be reported in the next chapter.

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 RESULTS

Figure 4.1 below shows the result of reference data collected from the T640 instrument between November 1st, 2022, and January 4th, 2023. Both the original data and the extracted data are presented in Appendix A &B. The sketch of the time series plots for the sensor data is presented in Figures 4.2 (a-c). The correlation graphs that describe the performance of the sensors are presented in Figures 4.3 (a-c). The generated AQI plot is presented in Figure 4.4.

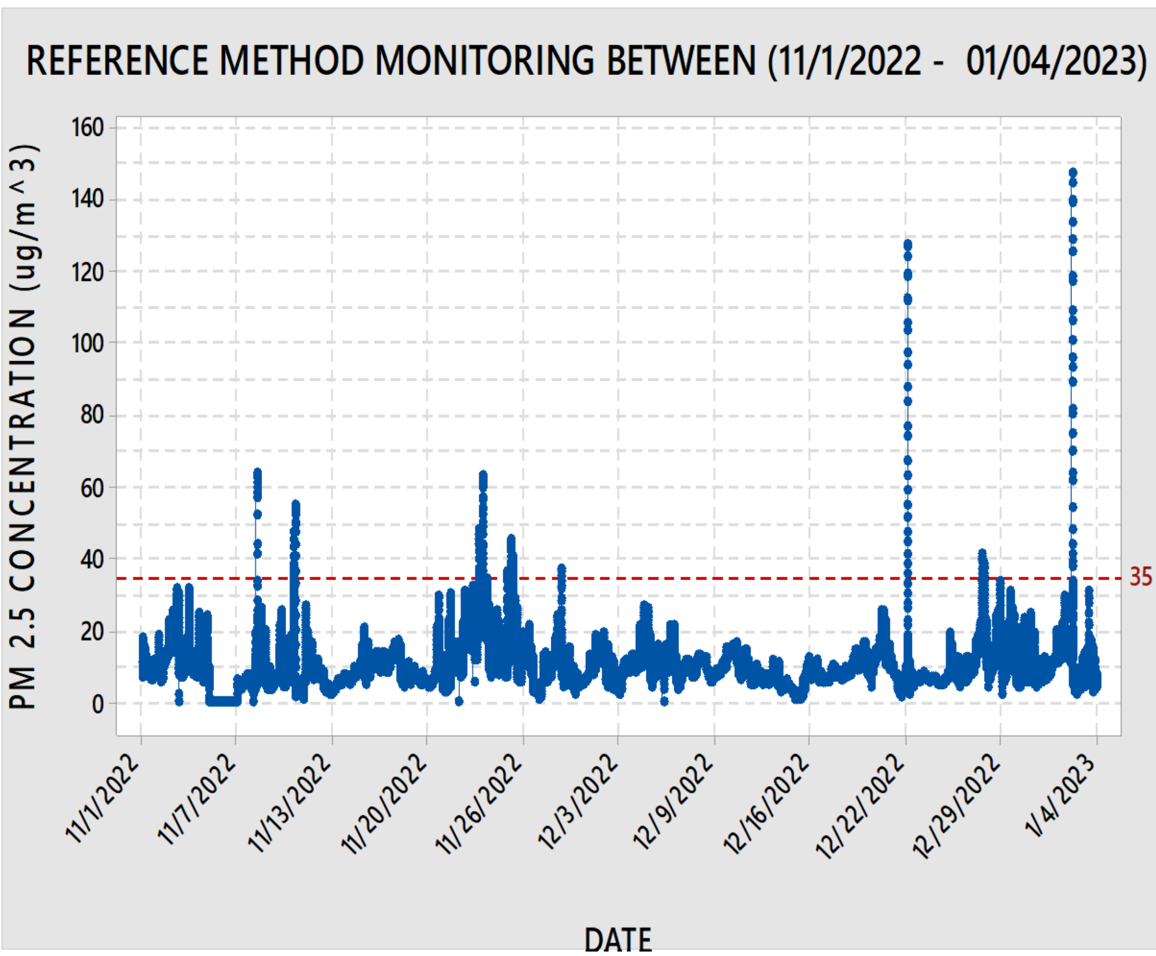


Figure 4.1: The plot of reference data collected at monitoring station.

4.1.1 DISCUSSION

The plot in Figure 4.1 shows atmospheric PM 2.5 data collected at North Braddock Air Quality Station between November 1st, 2022, and January 4th, 2023. The broken lines refer to the reference point (cut-off point). As shown on the graph, this reference point was exceeded in several numbers of days which means that a special atmospheric event is at play. Table 2-1 provides standards for all criteria pollutants and this table shows that the reference point has been exceeded. According to Table 2-1, the daily PM 2.5 concentration should not exceed 12.0 μg/m^3 for primary standard that protects our health, and 15.0 μg/m^3 should not be exceeded for the secondary standard that protects our resources. This daily PM 2.5 concentration provides a short time exposure limit and the potential health risk that may be associated with it. After careful observation of the historic data, it was observed an air inversion was at play during this period of data collection. Consequently, Allegheny County Residence with immunocompromised systems were advised to stay indoors.

4.2 SENSOR DATA PLOT

Figure 4.2 (a) below shows PM 2.5 data collected with PurpleAir Sensor (I) at North Braddock Air Quality Monitoring Station between November 1st, 2022, to January 4th, 2024. The plot shows a spike in measured atmospheric PM 2.5 concentrations. Although, this result correlates well with the reference data but the concentrations were overestimated which supported the general findings of low-cost sensors in literature.

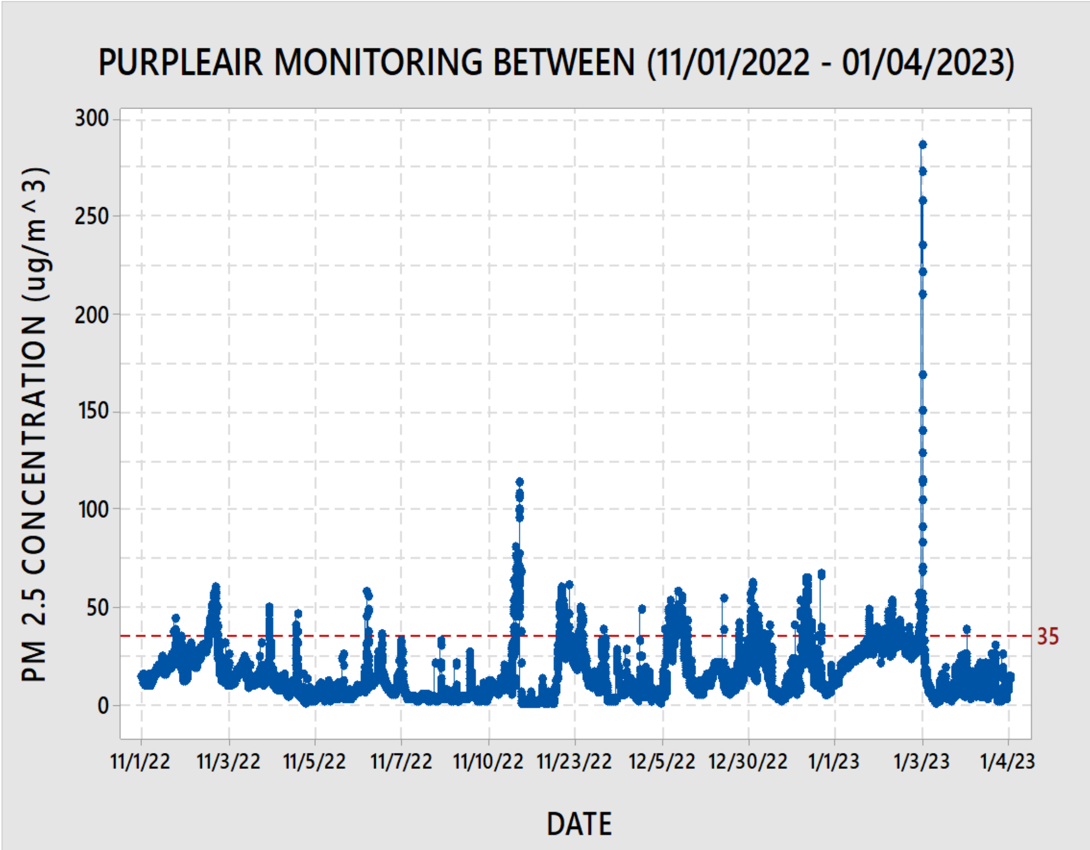


Figure 4.2 (a): PurpleAir Sensor (I) Plot

4.2.1 SENSOR (II)

Figure 4.2 (b) represents the data collected by PurpleAir Sensor (II) which also demonstrates spikes in measurements compared to the historical data.

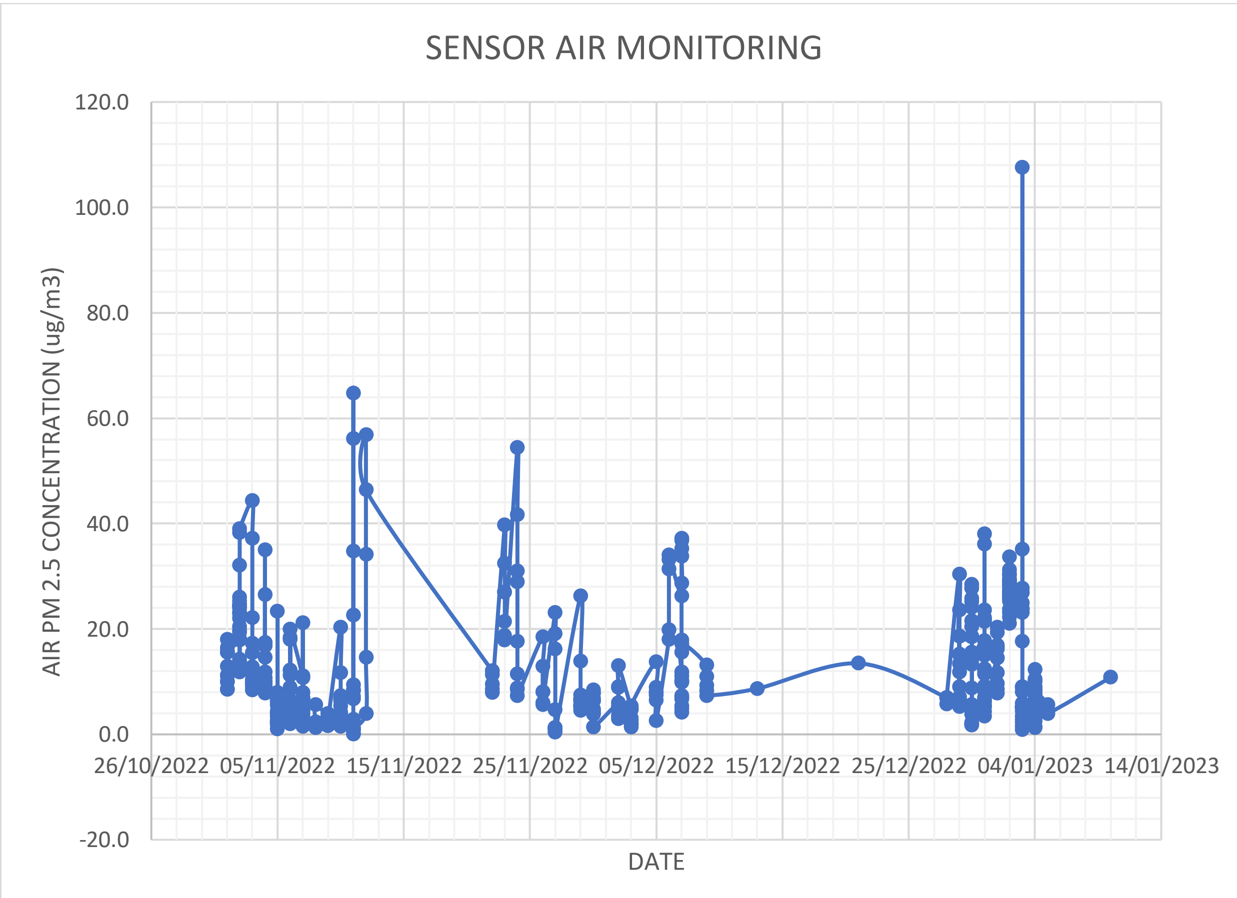


Figure 4.2 (b): PurpleAir Sensor (I) Plot

4.2.2 SENSOR (III)

Figure 4.2 (c) below shows PM 2.5 concentrations measured PurpleAir Sensor (III) at North Braddock Air Quality Monitoring Station. This sensor also demonstrates similar spikes in measurement as shown by other sensors.

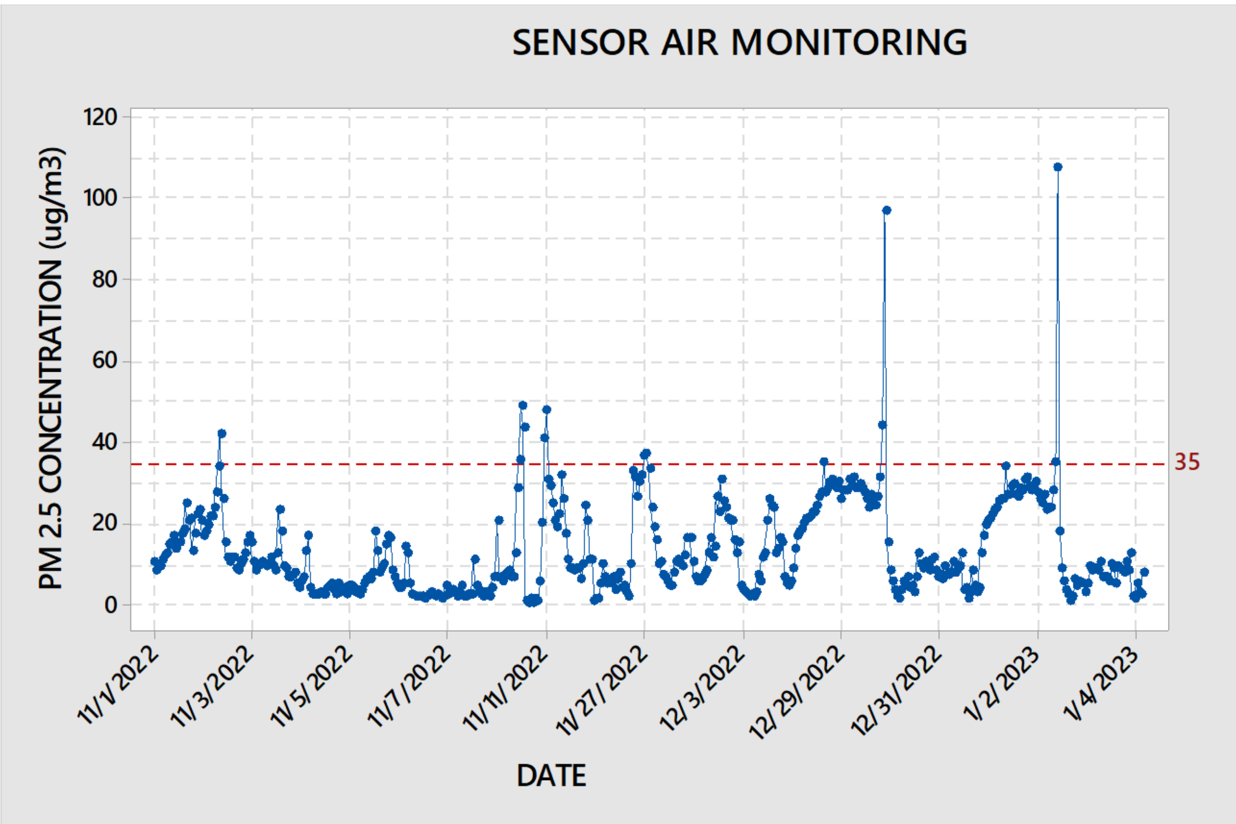


Figure 4.2 (c): PurpleAir Sensor (III) Plot

4.3 MATCHING OF REFERENCE AND SENCSOR DATA

According to section 3.4, to make sense of any collocated reference instrument and air sensor data monitors, we have to match the data acquisition timing of the two monitors (reference & monitors) together. The reference instrument report data hourly while the sensors report data by minutes therefore, the minutes sensor data were averaged into hourly data for each hourly reporting cycle of the reference instrument. During the analysis process, it was observed that the sensors did not report data for the entire monitoring period consequently; the corresponding missing data from the sensor was removed from the reference data to produce a matching result. Figure 4.3 (a-c) represents the results of this analysis. See Appendix (B-D) for the new table produced.

4.3.1 MATCHING DATA (REFERENCE & SENSOR 1)

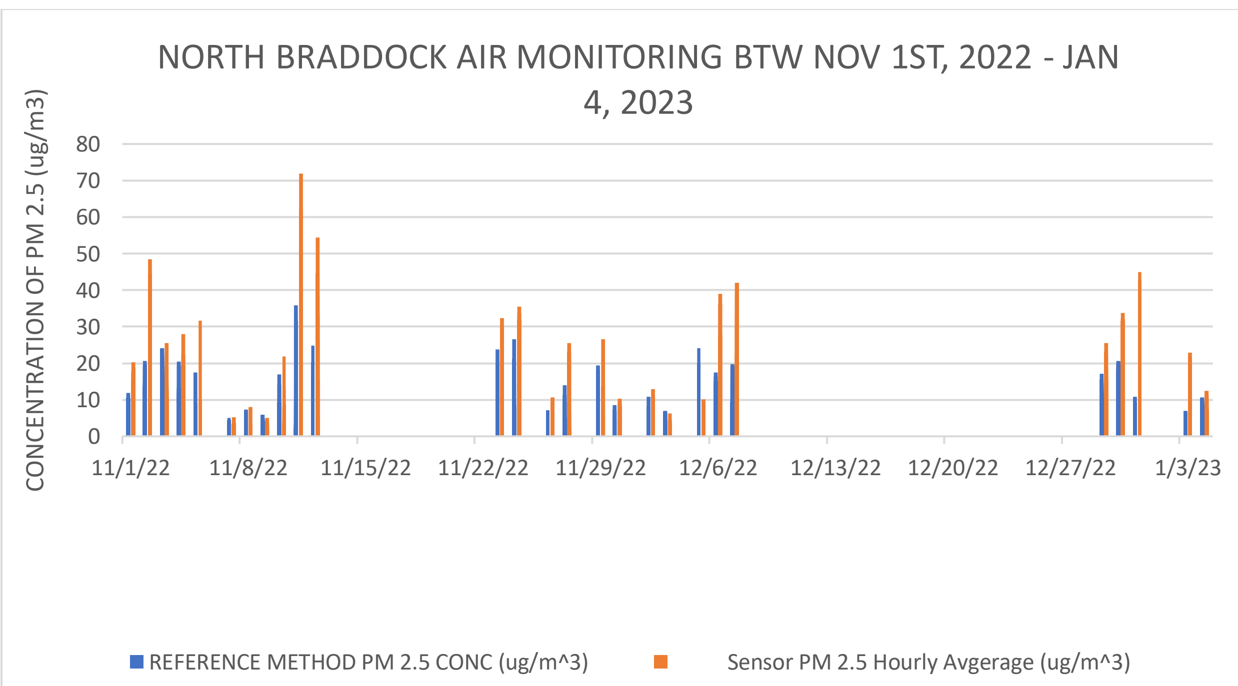


Figure 4.3 (a): Represent the relationship between reference and sensor 1 Data.

4.3.2 MATCHING DATA (REFERENCE & SENSOR II)

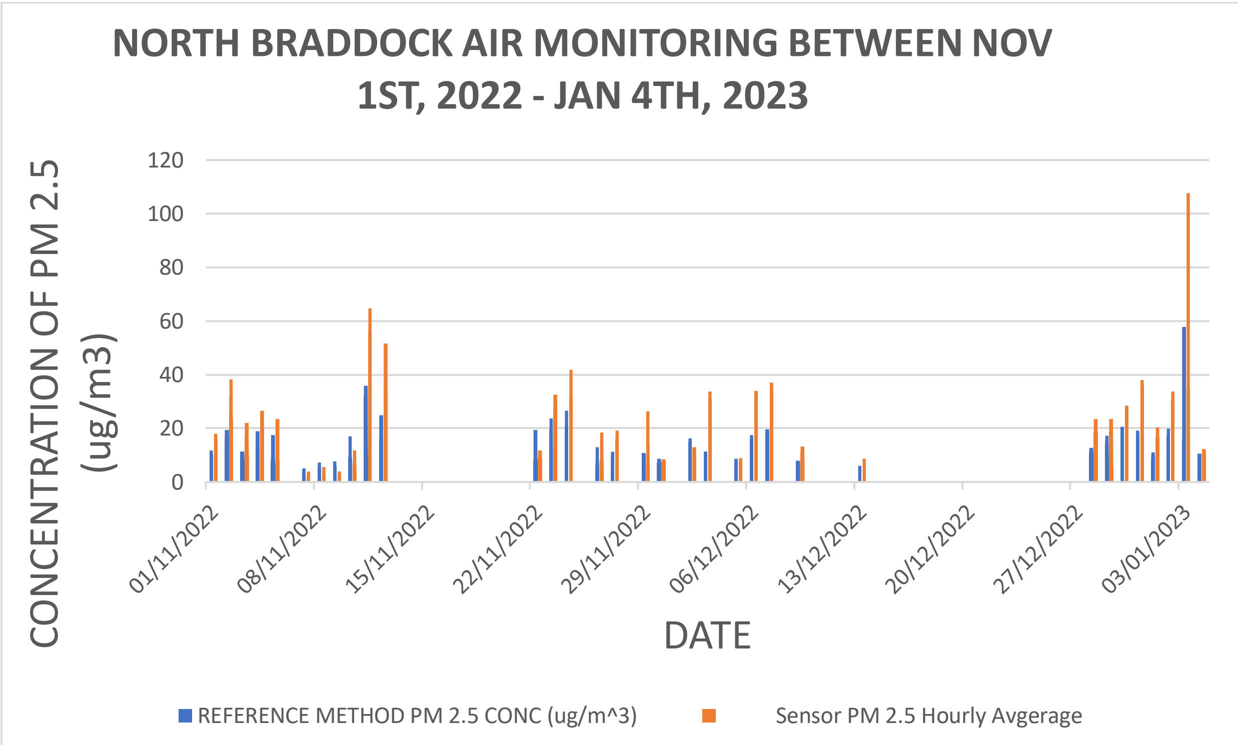


Figure 4.3 (b): Represent the relationship between reference and sensor 2 Data.

4.3.2 MATCHING DATA (REFERENCE & SENSOR III)

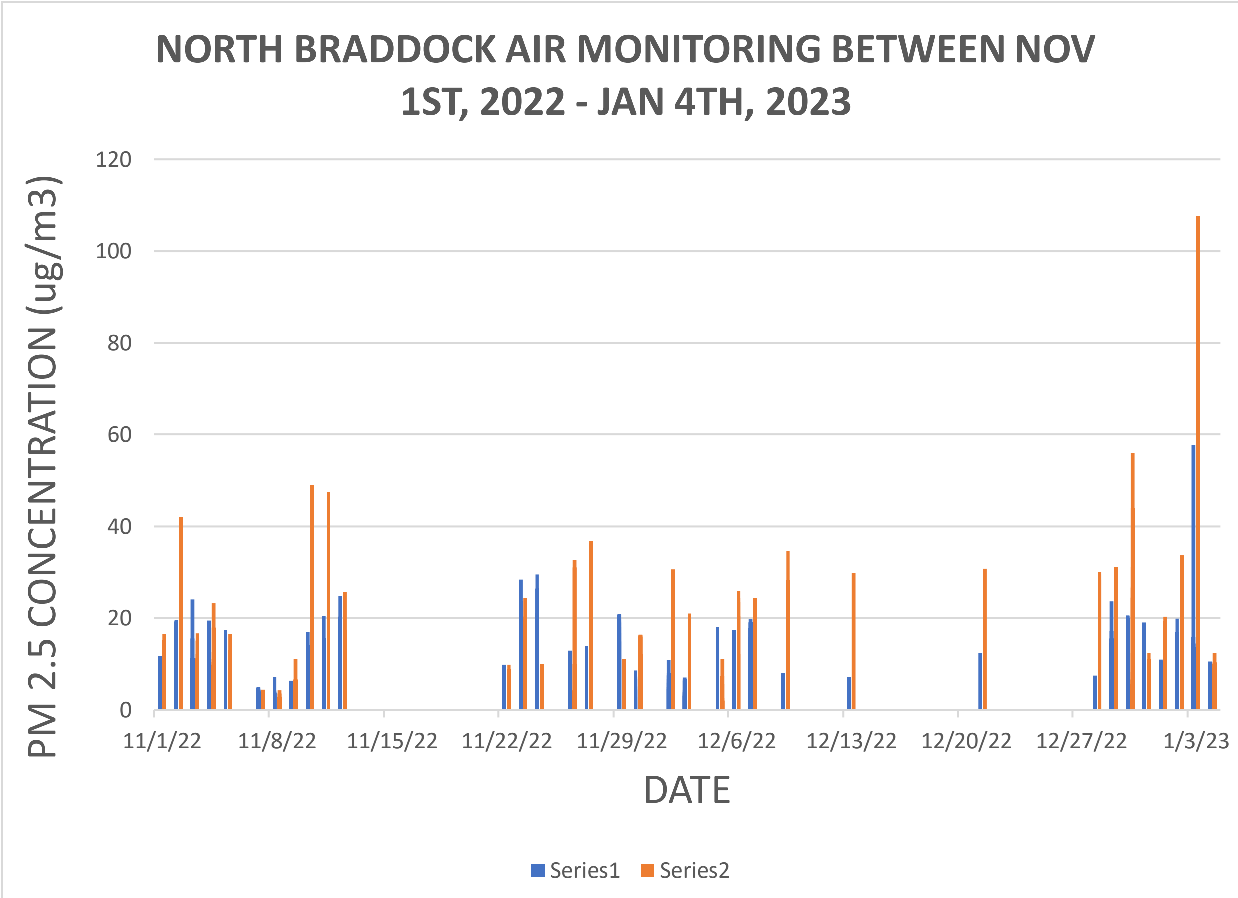


Figure 4.3 (c): Represent the relationship between reference and sensor 3 Data.

4.4 CORRELATION PLOTS BETWEEN PURPLEAIR SENSOR DATA AND REFERENCE DATA

4.4.1 SENSOR (I) VS REFERENCE DATA

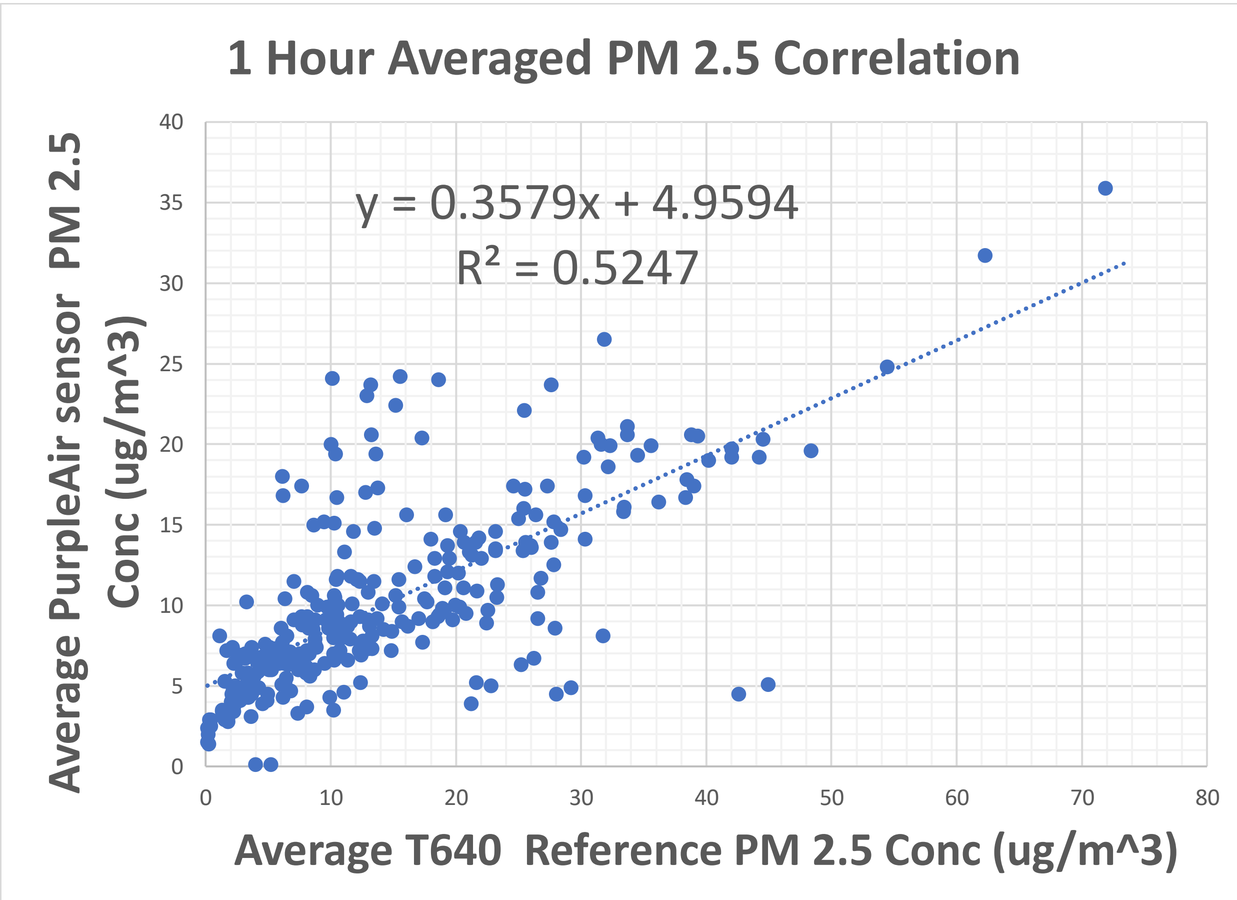


Figure 4.4 (a) represents the correlation between sensor I data and reference data.

According to section 3.4, the Environmental Protection Agency corresponds the relationships between sensor and reference instrument data to a Coefficient of Determination (R^2).

Where R^2 = 1, the closer the value of R is to 1, the more the sensor behaves like the reference instrument. Likewise, the closer the value of m is to 1, the more the sensor behaves like the reference instrument. Figure 4.4 (a) above represents the linear regression equation where the slope (m) = 0.36 and Coefficient of Determination (R^2) = 0.52, the intercept (b) = 4.96 (Biasness)

A perfect correlation between sensor and reference instruments should be 0.9 – 1.0 and a moderate correlation should be between 0.5 – 0.8. Since R^2 = 0.52, the correlation between the monitors is moderate.

Also, since most of the measurement taken by the sensor is elevated hence, this sensor may not be able to provide accurate measurement but might be able to help in hotspot identification as stated in the literature. Hence, this supports previous findings made by different researchers as stated in section 2.0.

4.4.2 SENSOR (II) VS REFERENCE DATA

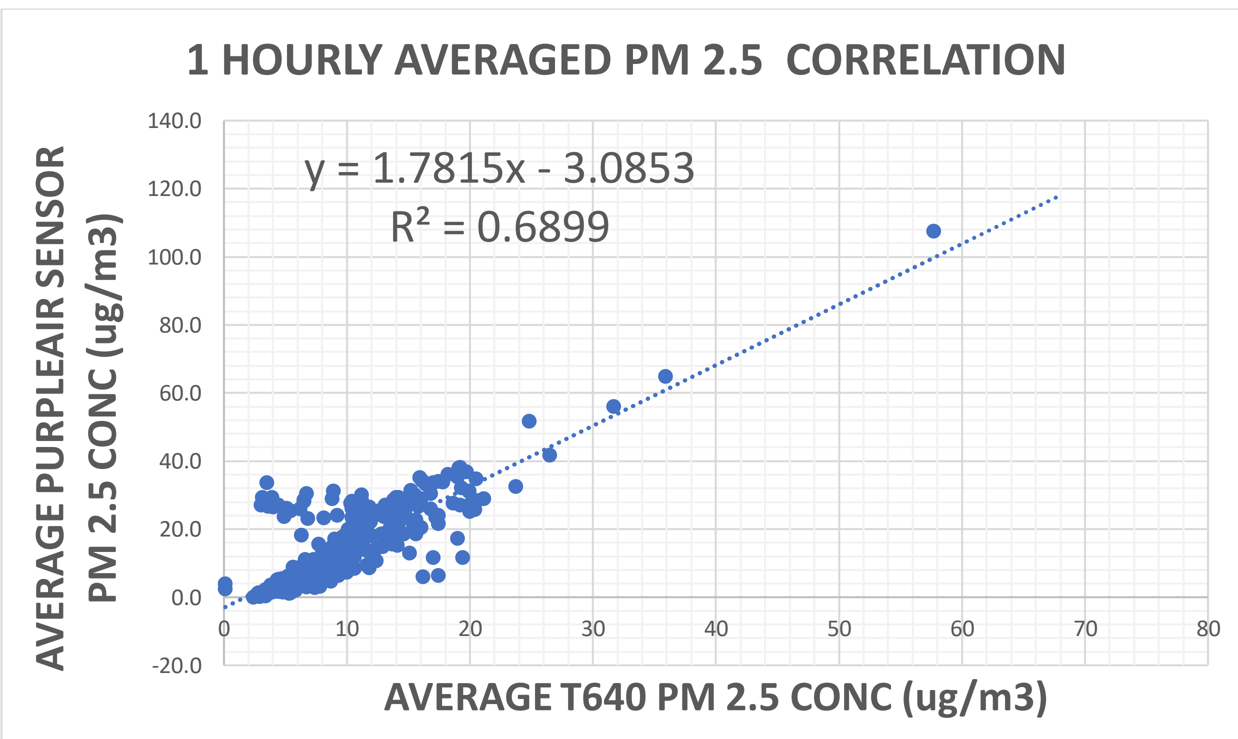


Figure 4.4 (b) represents correlation between sensor (II) & reference instrument.

It is shown from Figure 4.4 (b) that the correlation between the sensor and reference data is moderate being that $R^2 = 0.69$ and the slope (m) = 1.78, intercept (b) = 3.089 (biasness).

This simply means that the sensor can identify pollution hotspots, but the detected concentration may not be accurate hence, justifying the previous literature that the sensor may be helpful in locating pollution hotspots.

4.4.3 SENSOR (III) VS REFERENCE DATA

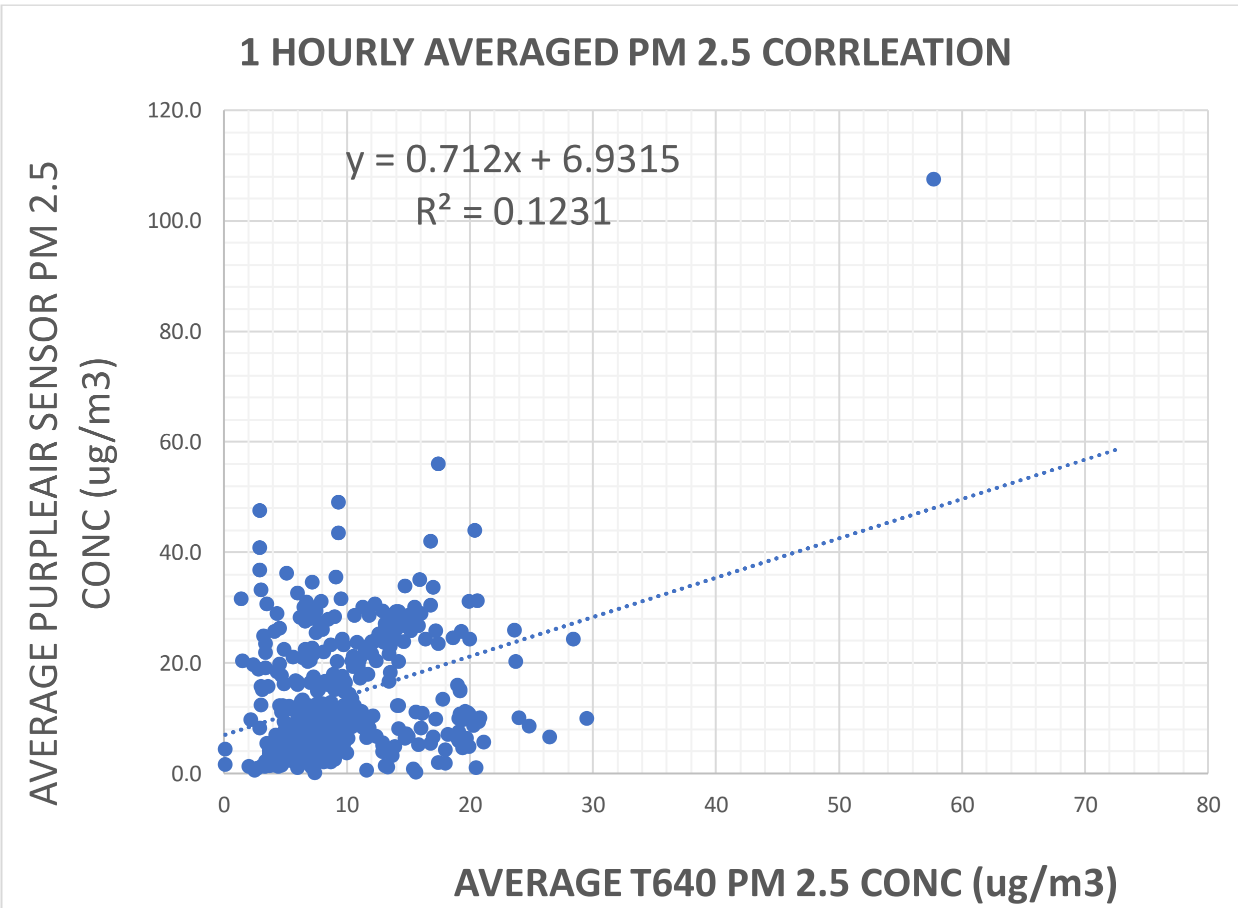


Figure 4.4 (c) represents correlation between sensor (III) & reference instrument.

Figure 4.4 (c) shows that $R^2 = 0.12$, m = 0.71, and intercept (b) = 6.93 (biasness).

R^2 must be closer to 1 to establish a perfect relationship between the sensor and the reference instrument. Based on Figure 4.4 (c) above $R^2 = 0.12$ which is far from 1.0 therefore, this linear model did not establish any relationships between the two monitors. This low value of R2 means that only 12 % of the variability in the data can be explained by the regression model while 88 % can be accounted for by other factors which cannot be explained by this model. This shows that the model is not efficient in establishing the projected relationship. Further analysis may be needed to explain these other factors that cannot be described with this model. By considering these other factors, the data may be represented more accurately. According to the literature, even sensors of the same model may not generate the same reading even after being corrected as stated in section 2.0.

4.5 CORRECTED SENSOR DATA

According to Figure 4.4 (a-c), R2 = 0.52, 0.69 and 0.12.

According to figure 4.5 (a-c), R2 = 0.58, 0.67 and 0.09.

Comparing Figure 4.4 (a-c) and Figure 4.5 (a-c), it will be observed that there was no considerable improvement in the Coefficient of Determination. According to the literature, sensor monitors are often affected by environmental factors such as temperature and relative humidity (RH). Relative humidity was more documented in the literature that it could negatively influence sensor performance due to water uptake hence, correction may and may not make a significant improvement to sensor data.

4.5.1 SENSOR (I) VS REFERENCE DATA

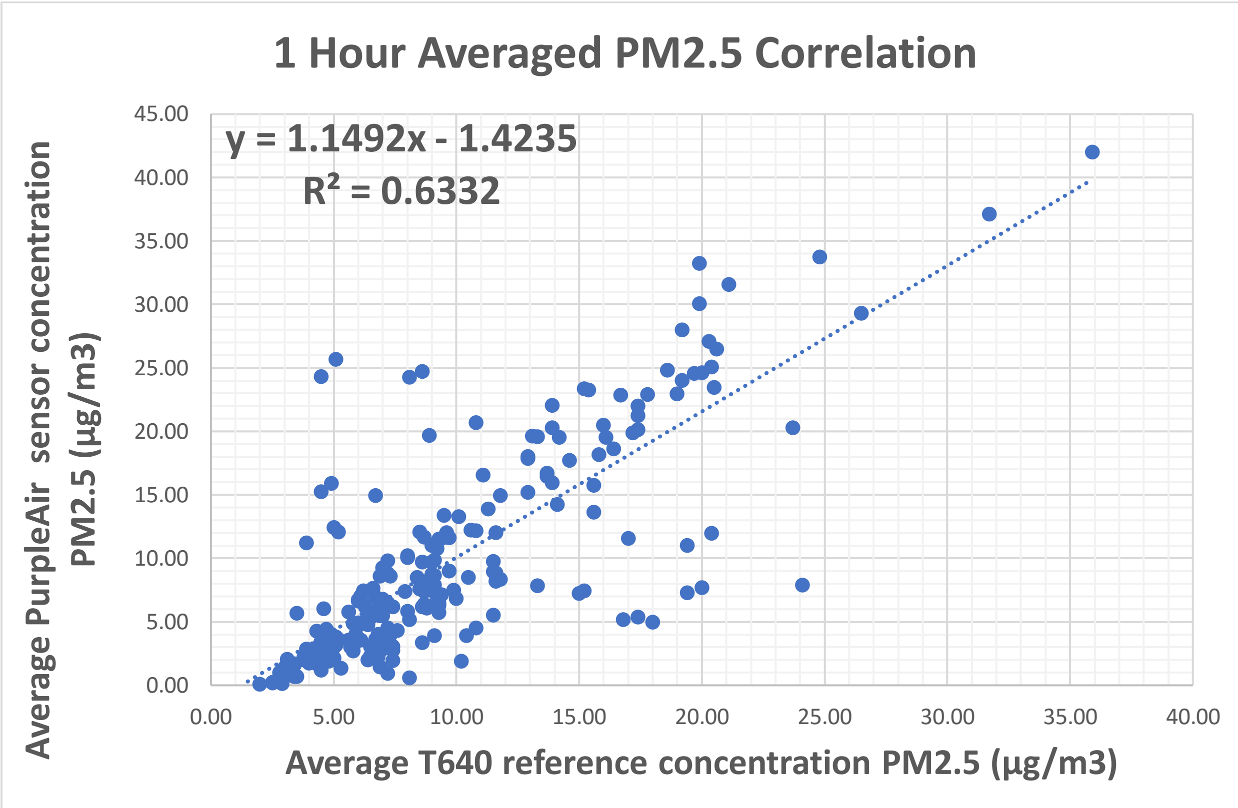


Figure 4.5 (a) represents corrected sensor (I) data vs reference data.

4.5.2 SENSOR (II) VS REFERENCE DATA

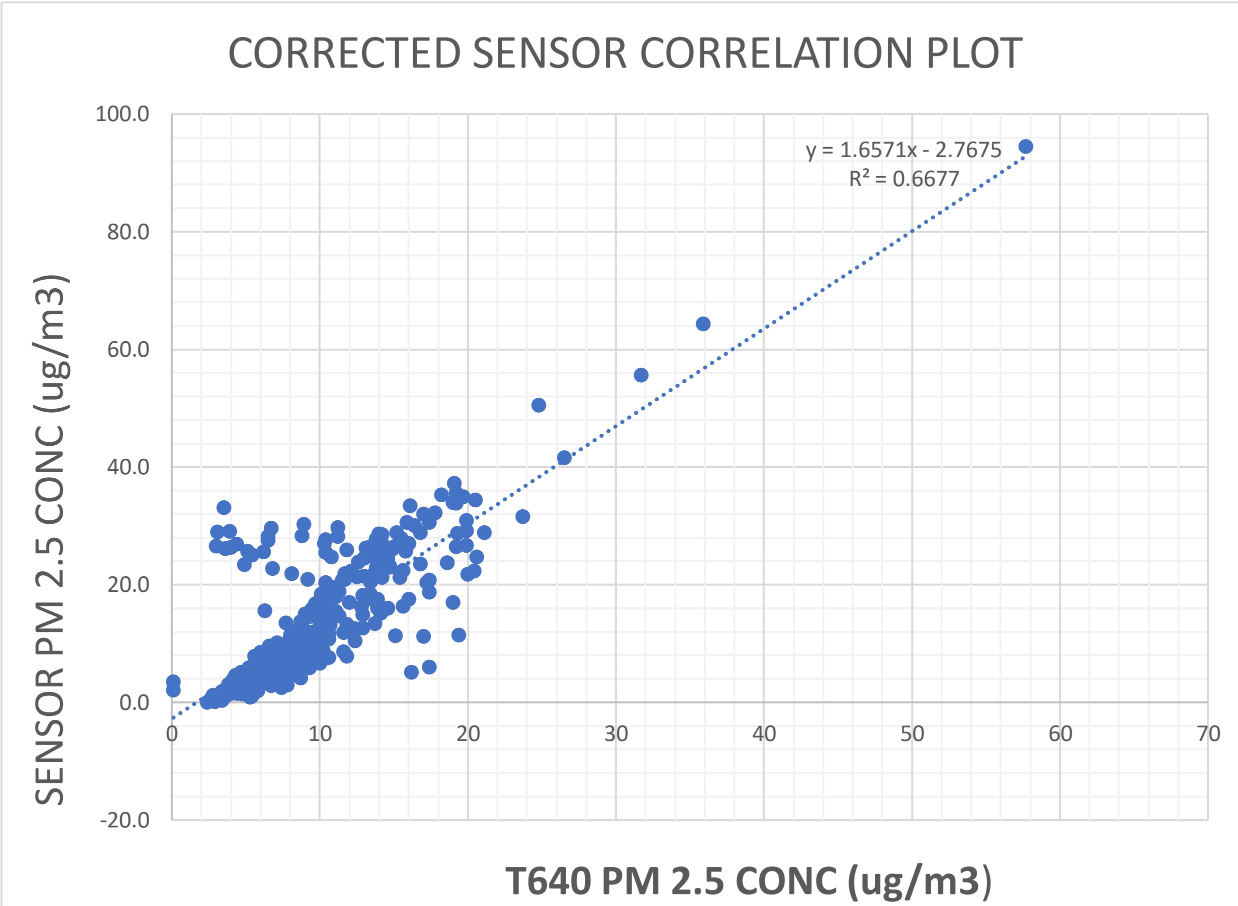


Figure 4.5 (b) represents corrected sensor (II) data vs reference data.

4.5.3 SENSOR (III) VS REFERENCE DATA

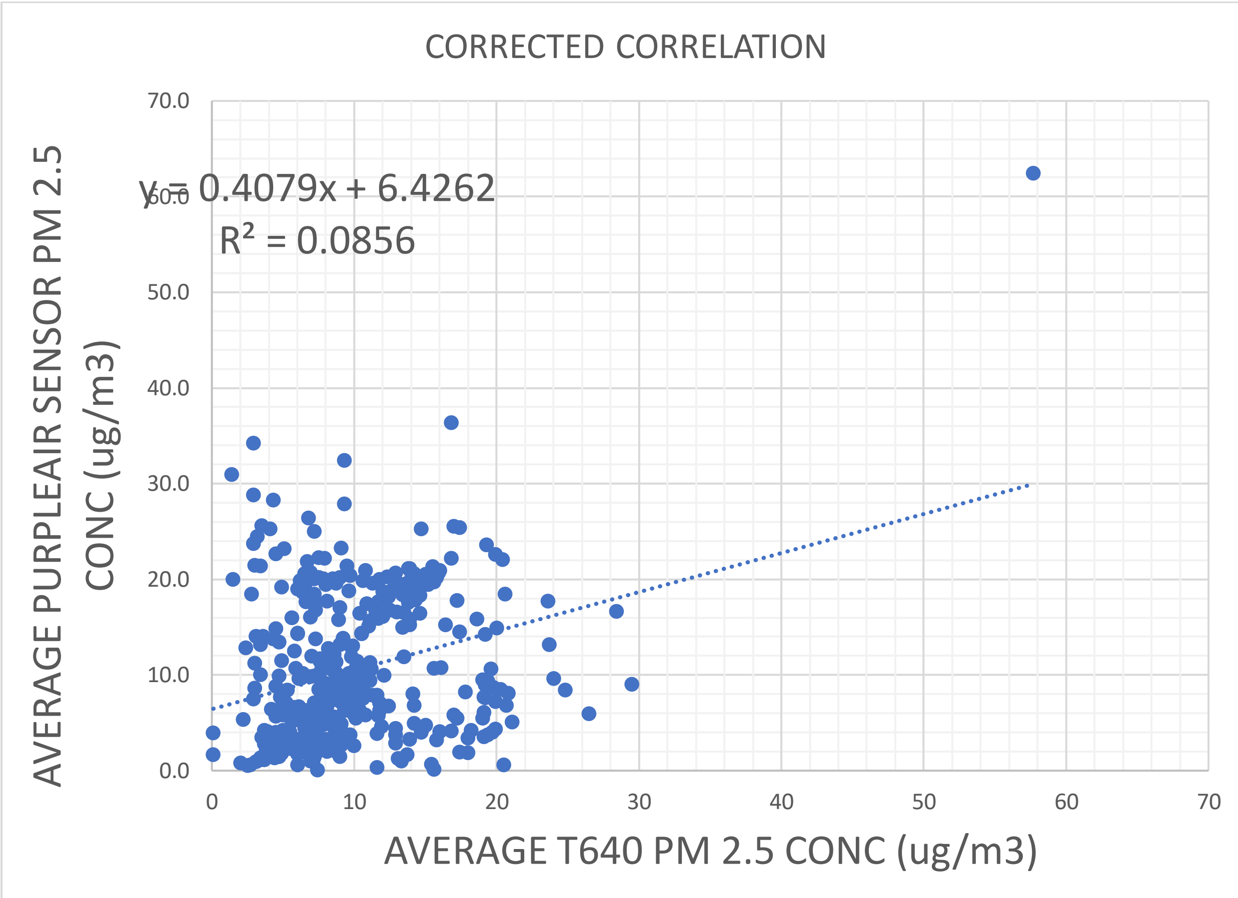


Figure 4.5 (c) represents corrected sensor (III) data vs reference data.

4.6 General Discussion

Based on this project, two out of the three sensors used for this work demonstrated a moderate relationship with the reference instrument. Research shows that low-cost sensors have several limitations compared to monitoring instruments. Relative humidity and temperature were mostly accounted for in the literature as the most common environmental factors that affect readings collected by air sensors. But since relative humidity is the most ranked among the environmental factors, the effect of RH was removed using an aerosol equation demonstrated by Malings *et al*. (2019). This equation was used because calibration made in one place may not translate into another place. While reference instruments have been designed to flag any unusual irregularities thus, effective performances have been attained for reference instruments. According to this work, calibrating the sensors does not guarantee improvement in the data quality for example, removing the effect of RH, only makes a little bit of improvement to sensor one, no changes to sensor 2, and performance decreased for sensor 3. These sensors are of the same model with the expectation of getting the same results. According to the EPA, the use of three sensors is to be able to detect bad sensors. A bad sensor will depict a low R^2 value. Since sensor 3 has a very low R^2 value, we can either conclude that the sensor is bad, or it is experiencing early degradation which should probably occur within 2-3 years of deployment, or it is probably affected by winter conditions during which the data was collected. Since a concrete conclusion cannot be made as to why the third sensor is showing very low performance, these sensors' performances need to be re-evaluated in another season for comparison therefore, more work needs to be done in the future. It is important to note that since sensors were detecting the presence of PM 2.5 even though, at a high concentration compared to the reference instruments, it is, therefore, recommended to use sensors for community monitoring since reference monitors cannot be installed in every community due to the huge cost involved with purchasing reference instruments.

4.7 HEAT MAP OF THE STUDIED AREA USING AQI

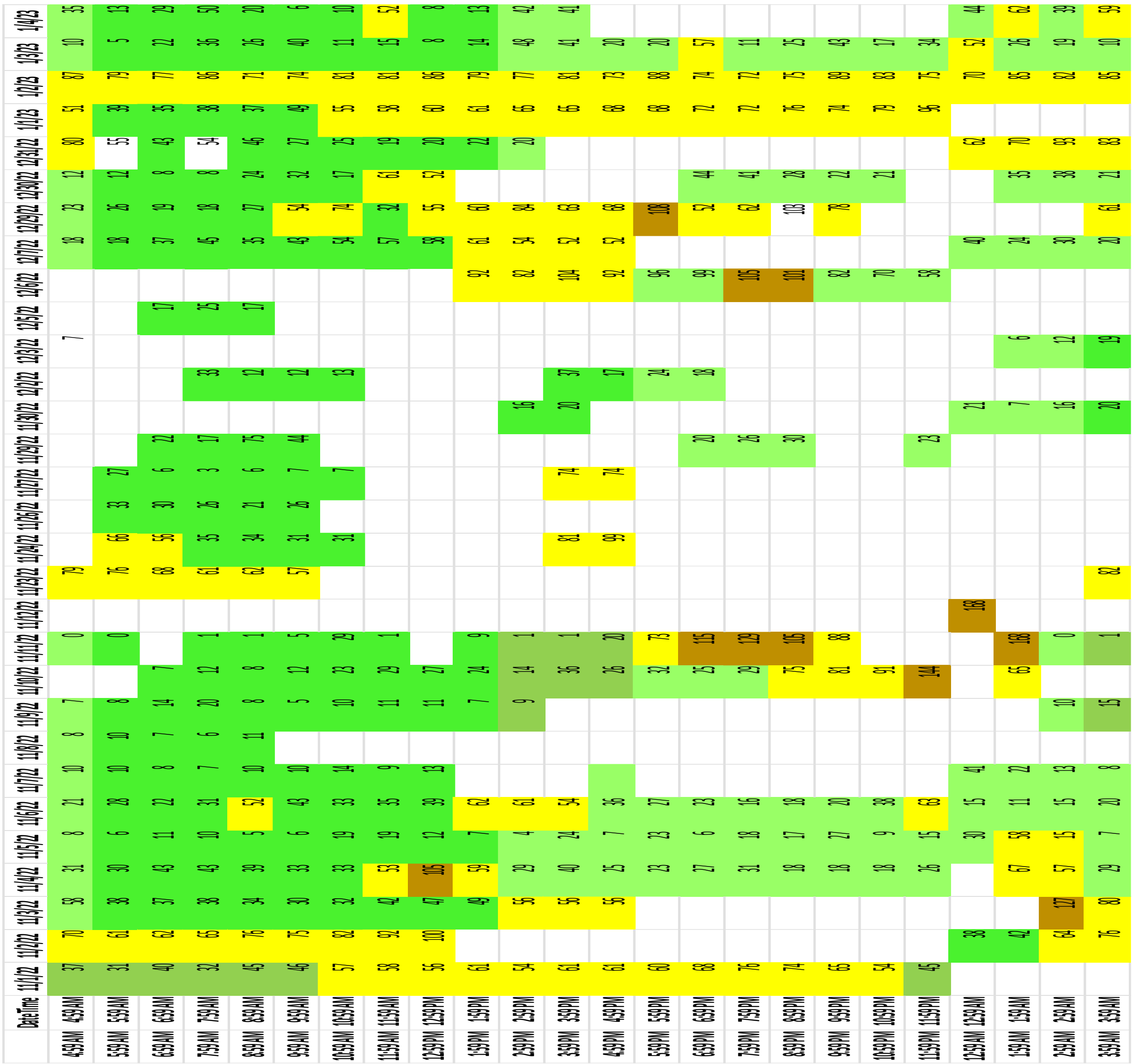


Figure 4.7 (a) represents heat map of the studied area using AQI.

4.7.1 AQI OF STUDIED AREA

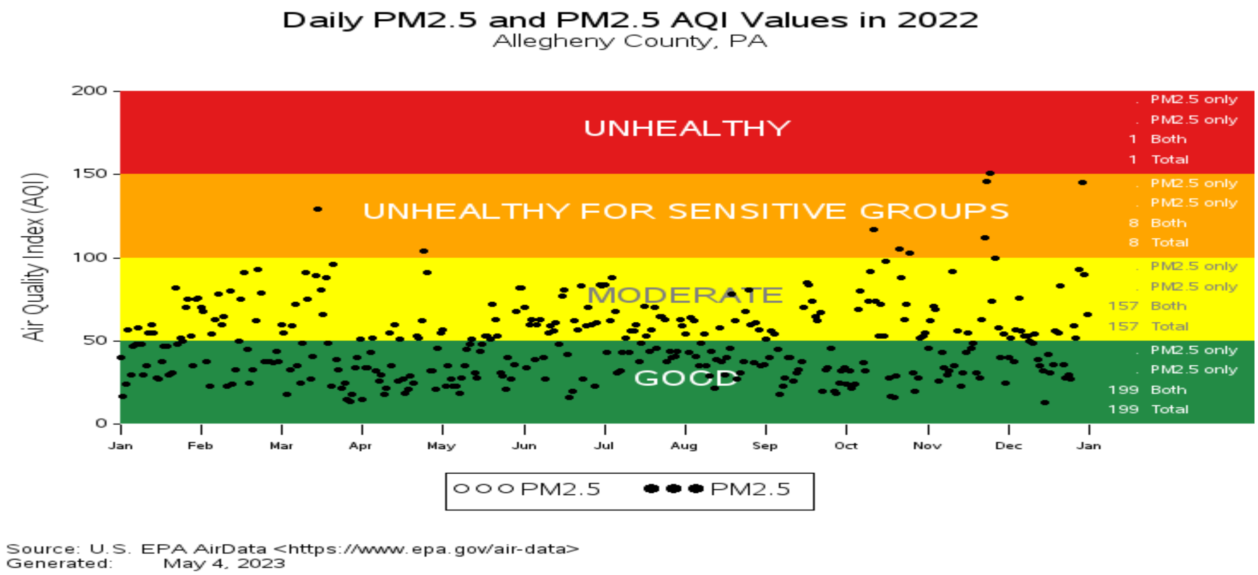


Figure 4.7 (b) represents the AQI of the studied area from EPA website.

The Air Quality Index (AQI) acts as an important tool aimed at promoting public health, educating people about air pollution problems, directing environmental policy, and ensuring effective communication of air pollution concerns to the public. It contributes to a healthier living environment while demanding every individual to take action about protecting themselves against the impact of air pollution problems. For this research, the AQI category falls within good (green), moderate (yellow), and unhealthy for sensitive groups (brown). Therefore, people with immunocompromised systems were advised to stay indoors. The AQI heat map generated manually correlates well with the AQI map of the studied area downloaded from the Air Quality System (AQS) website.

5.0 CONCLUSION

Having considered three different PurpleAir sensors data with T640 reference data, it was observed that the first-second sensors data established moderate relationships with the reference data. In contrast, the third sensor cannot establish any relationship with the reference instrument with $R^2 = 0.12$, which means the linear model can only account for 12 % of the variability in the data and another factor is responsible for the remaining 88 %. EPA recommends the use of three sensors so that we can recognize a bad sensor with a low coefficient of determination (R^2). A chat with Air Quality Scientist from ACHD stated that there are 10-15% of broken sensors out of the box. Because air sensor performance can be impaired by environmental factors such as temperature and relative humidity, therefore, a correction was made to remove the effect of RH since it is commonly accounted for in the literature using an equation established by Malings *et a*l. (2019). Correcting relative humidity to improve sensor data analysis did not make any significant changes to the data, but the sensors could locate pollution hotspots. Since we cannot determine the exact factor responsible for the poor performance of the third sensor, it is necessary to obtain more data for these three sensors to re-evaluate their performance in another season because we can't tell if the third sensor was affected by the coldness of the winter season or it is a bad sensor with low R^2 value or it is experiencing early degradation in performance which should happen after a year or between 2-3 years after deployment. Therefore, to accurately determine the interpretation of these sensors, more analysis is needed for future work. Furthermore, this project identifies the potential of air sensors for community monitoring since they can detect the same parameters as reference instruments, only at elevated levels. Since it is impossible to install reference instruments at every location, air sensors may be used as alternatives considering the affordable cost.

REFERENCES

Austin, E., Novosselov, I., Seto, E., & Yost, M. G. (2015). Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor. PLoS One, 10, e0137789.

Byrnes, M. (2012, June 5). What Pittsburgh Looked Like When It Decided It Had a Pollution Problem. Bloomberg. Retrieved from https://www.bloomberg.com/news/articles/2012-06-05/what-pittsburgh-looked-like-when-it-decided-it-had-a-pollution-problem.

Bulot, F. M. J., Johnston, S. J., Basford, P. J., Easton, N. H. C., Apetroaie-Cristea, M., Foster, G. L., Morris, A. K. R., Cox, S. J., &Loxham, M. (2019). Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. Scientific Reports, 9, 7497. https://doi.org/10.1038/s41598-019-43716-3.

Chakrabarti, B., Fine, P.M., Delfino, R., Sioutas, C. (2004). Performance evaluation of the active-flow personal DataRAM PM2.5 mass monitor (Thermo Anderson pDR-1200) designed for continuous personal exposure measurements. *Atmosphere Environment*. 38: 3329–3340.

https://doi.org/10.1016/j.atmosenv.2004.03.058

U.S. Environmental Protection Agency. (2017). Quality Assurance Handbook for Air Pollution Measurement Systems, Volume II, Ambient Air Quality Monitoring Program (EPA-454/B-17-001). epa.gov/sites/default/files/2020-10/documents/final_handbook_document_1_17.pdf

Environmental Protection Agency. (2018). How to evaluate low-cost sensors by collocation with federal reference method monitors: EPA report. Office of Research and Development, National Exposure Research laboratory. EPA/600/R-18/026 https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=343236

Environmental Protection Agency. (2018). Teledyne T640 Continuous PM2.5 Manual. Retrieved from https://www.epa.gov/sites/default/files/2018-03/documents/t640x_pm2.5_manual_0.pdf

Environmental Protection Agency. (2022). List of Designated Reference and Equivalent Methods. https://www.epa.gov/ttn/amtic/criteria.html (40 CFR Part 53)

Environmental Protection Agency. (2022). The Enhanced Air Sensor Guidebook. Center for Environmental Measurement and Modeling Office of Research and Development, CA. https://www.epa.gov/air-sensor-toolbox. EPA/600/R-22/213

Gao, M., Cao, J., &Seto, E. (2015). A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. Environmental Pollution, 199, 56-65.

Jessica, T., L'Orange, C., Mehaffy, J., Miller-Lionberg, D., Hofstetter, J. C., Wilson, A., &Volckens, J. (2020). Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers. *Atmospheric Environment*, 220, 117067.

Lance W., Jianzhao Bi., Wayne R. Ott., Jeremy S., Yang Liu. (2021). Calibration of Low-cost PurpleAir Outdoor Monitors using an improved method of calculating PM2.5. *Atmosphere Environment*, 256.

Liu, D., Zhang, Q., Jiang, Chen, D.-R. (2017). Performance Calibration of Low-cost and Portable Particulate Matter (PM). *Journal of Aerosol Science*, 112, 1-10.

Nelson, K. (2016). Dirty Air Monitors: Are they telling us the whole truth? [Blog post]. Retrieved from https://www.ucsusa.org/resources/dirty-air-monitors-are-they-telling-us-whole-truth

Wallace, L. (2022). Intercomparison of PurpleAir sensor performance over three years indoors and outdoors at a home: bias, precision, and limit of detection using an improved algorithm for calculating PM2.5. *National Library of Medicine*, 22(7):2755.