SENSOR NETWORK FOR AIR POLLUTION MONITORING

by

Ashlin Saju
B.Tech, Mahatma Gandhi University, Kerala, India, 2016

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS OF THE DEGREE OF
MASTER OF SCIENCE
IN
COMPUTER SCIENCE

UNIVERSITY OF NORTHERN BRITISH COLUMBIA
September 2019

© Ashlin Saju, 2019

Abstract

Needs to be filled later

Contents

Abstract ii

Table of Contents iii

List of Tables v

List of Figures vi

Acknowledgement vii

1 Introduction **1**

2 Literature Survey **2**

3 Design and Implementation of Air pollution Monitoring System **3**

3.1 Design Goals 3

3.2 Targeted Pollutants 5

3.3 System Architecture 6

3.3.1 Hardware Architecture 7

3.3.2 Sensor 9

3.3.3 Communication Middleware 12

3.4 System Overview 13

3.5 IoT Architecture 13

4 Calibration Of Low Cost Sensor **15**

4.1 Calibration Procedure 17

4.2 Macro Analysis Tool -MAT 18

4.3 Linear Regression 20
Bibliography 22

List of Tables

3.1 Technical specification 9

List of Figures

3.1 System Architecture 7

3.2 Arduino Ethernet Board 8

3.3 Sensitivity characteristics curve [1] 11

3.4 System Overview 14

3.5 Printed Circuit Board 14

4.1 MAT Output 18

4.2 Correlation Graph 19

4.3 Time-Series Graph 19

4.4 Linear Regression Model [2] 20

Acknowledgement

I would like to express my deep sense of gratitude to my worthy supervisor Dr. Alex Aravind for allowing me to carry out my work under his invaluable guidance. His interest in scientific problems and moral inspiration have made current work possible. I am extremely indebted to the Chairperson, Department of Computer Science, Northern University of British Columbia, Prince George, for providing all the necessary facilities to complete this work. Also, a special thanks is extended to Dr. Stephen Rader and Dr. George Jones for offering their valuable assistance as my committee members.

I am also indebted to my labmates Apratim, Ashlin, Conan, Brae, Davis, Suresh, Shanthini, Darshik, Rahim, Ketan, Arthi, Raja and Rodrigo for their support, discussion, encouragement, and pleasant company. Special thanks to the teaching and nonteaching staff members, UNBC, for their help in all possible ways.

It is a pleasure to acknowledge the good company of fellow researchers and my friends Hardev, Gurjeet, Ashish, Ramneesh, Johny, Priyanka, Amit, Navjot, Bashet, Dorob, Bryan and many others.

Finally, I want to express my deepest gratitude to my family for their love, support and encouragement, without which this work would not have taken this shape.

Prince George, BC
September , 2019 *(Ashlin Saju)*

Dedicated

to

My Parents

Chapter 1

Introduction

Chapter 2

Literature Survey

Chapter 3

Design and Implementation of Air pollution Monitoring System

The progress in technology has made it easy for researchers to explore more in different areas. One such development was is in the area of sensor technology which completely changed the outlook of different application level problems. In this chapter, I share my hands-on experience in the development, design, integration, and operation of the air pollution system using commodity sensor. Earlier, the approach for understanding air pollution used complex and stationary equipment which collects data and used these data for analyzing, but things have changed after the low cost, easy to use, portable sensors came in markets [3].

3.1 Design Goals

There are many factors which need to be considered for the development of a simple yet reliable system. In this section, we have mentioned the factors which should be considered for an effective air pollution monitoring system.

1. Sensor Identification

The very first task is to figure out which all sensors need to be included for the completion of the system. There are sensors available in the market for the measurement of almost all types of gases in the atmosphere. It should be very clear that which all gases need to be measured and this definitely changes from region to region as in certain places the concentration of a particular gas is more. Having said that, there will be a certain set of gases which must be included for measurement regardless of the region.

2. Communication Module

   As the system is completely based on wireless sensors the selection of data transmission is another crucial factor. The communication between the server and the sensors should be taken into consideration. The collected data from the sensors should be transferred over a database or to the sever. For that the type of communication module can be either Wi-Fi or bluetooth module.

3. Reliability

   The success of the system depends upon how much accurate the data is. The value which we obtain from the sensor should make sense to the audience. There will be a lot of noise coming with the collection of data, the sensor should have the ability to remove the noise data or it should allow the programmer to make changes or apply certain algorithm so that the data sets will be refined.

4. Easy Integration

   The integration of sensors with the processor is one important factor that needs to be kept in mind. Some sensors can be easily integrated with any processor but others needs driver codes to be written in order to work with the processor.

5. Printed Circuit Board

   The final system should be build on a printed circuit board as it is more dependable. Circuit build on basic breadboard might even come out as it is not permanently fixed

and this will cause frequent breakdown. Its always easy to work on breadboard but that will be useful only for the initial set up. The system should be transformed to PCB.

6. Maintenance

   In case of any sensor damage it should be easily replaceable which means the complete system should be a plug and play type model. On building up such a model like that will help in debugging the problems caused by sensors if any. It should also be considered that the sensors selected for the system should be easily available in market so that it can be replaced if needed.

7. Easy Replication The idea behind creating such a system is that it can be replicated by anyone without even knowing the dept knowledge. The system should be designed in such a way that it should use the most available sensors and processors in the market. The programming part of the sensors to processor will be easy if the selection of processor is simple. This could definitely bring down a lot of work done at the hardware level.

8. Low Cost

   Within the available sensors in the market one could find sensors ranging from a very low price to costliest of all. There was a budget set for the the complete system and finding the right sensors with the affordable cost is one crucial factor.

3.2 Targeted Pollutants

Our surrounding is filled with various gases, these gases will become harmful if the concentration of it increases to an undesired level. On the development of a air pollution system measurement of all the gases in the atmosphere is not necessary as the collected data from all the sensor will make no sense to the public. Our main idea here is to make the general people

aware about the dominant gases and the extend of health hazard caused by these gases. This can be identified through different indexes know as Air Quality Health Index(AQHI) which is a scale from one to ten developed by health and environmental professionals [4] and Air Quality index (AQI)which gives the level of air quality status in an area [5].

The development of such indexes by the scientists will give the general public more idea of the pollution. The main gases to be included for the measurement for the indexes are $PM_{2.5}, O_3, NO_2$, and CO along with temperature and humidity sensor for awareness. These gases are mainly caused due to industrialization, urbanization and motorization [6]. Industrial and vehicles release greenhouse emissions which are largely responsible for air pollution [7]. The sensors thus can be limited to five which will also make the system compact.

3.3 System Architecture

In this section we describe the architecture for air pollution monitoring system which include a hardware side and a software side.The system is designed in such a way that it collects the data through the sensors, performs certain mathematical equation on the collected data and calculate the indexes then transfer these data to an IoT platform where it is visualized. The hardware section includes multiple sensors, processor and also on the wireless communication module for transmitting and receiving signals. Second, we will discuss the IoT side which includes the visualization part. The complete idea of the system is as shown in the figure and each part along with the sensor specification, implementation, design will be discussed further in the section.

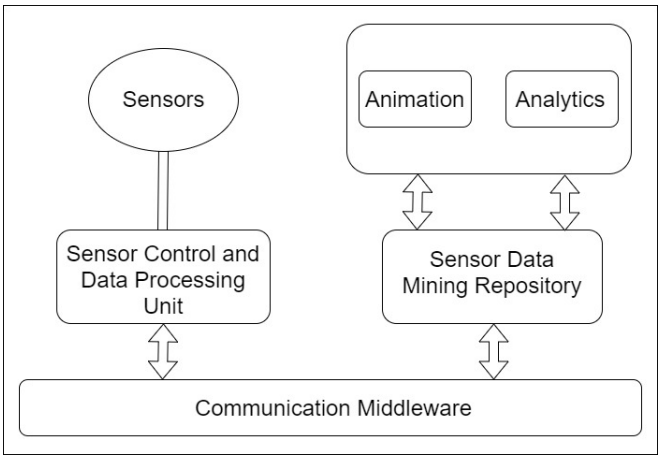


Figure 3.1: System Architecture

3.3.1 Hardware Architecture

Sensor Control and Data Processing Unit

This unit is the core for the pollution monitoring as it is where all the other modules are connected including the communication middleware. This module does the following function:

1. Control the sensors in collecting data.
2. It filters and processes the collected data and forward to the software.
3. Provide the necessary voltage for all the hardware connected to it.

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

Arduino is an open source physical computing platform that is divided into two parts, one is the hardware which is the board itself in which the external components are added to and other is the software which is the development environment for the processing language. It is very easy and simple to use the board with any 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 [8]:

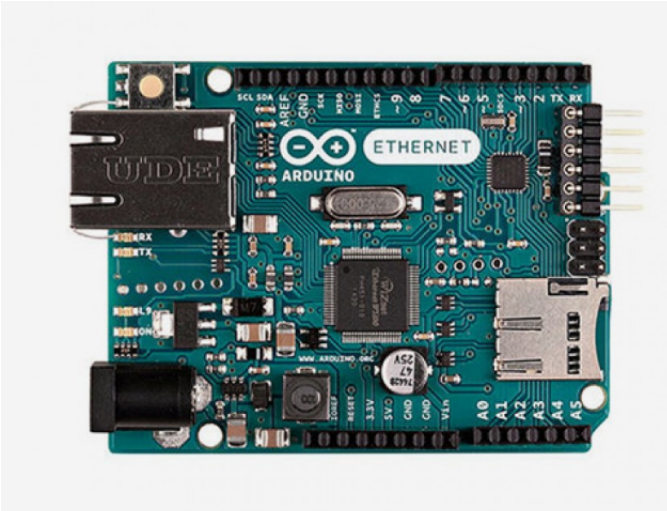


Figure 3.2: Arduino Ethernet Board

1. Arduino is multi-platform and can be used with windows,mac and linux.

2. The programming is done via USB cable and not serial port and is useful as modern computers don't have serial port.

3. As it is open source hardware and software all the necessity for an external hardware to be worked on like circuit diagram, the code can be easily downloaded.

4. There is an Integrated Development Environment (IDE) which can be used as an interface for talking with the hardware and is very simple.

5. There is an active arduino forum in which many researchers or developers who are working on projects contribute their ideas and will help in trouble shooting.

6. The cost of the hardware is very cheap and will come under 60 CAD and is easily affordable.

The board can be powered either by using a FTDI cable/USB serial connector, external power supply or using an optical Power Over ethernet module (PoE). The ATmega328 has 2 KB of SRAM and 1 KB of EEPROM. The detailed specification [9] of the board is given below in the table.

Table 3.1: Technical specification

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

Having all these specification gives the freedom for researchers to explore with different electronic devices easily. The board is programmed with the help of arduino programming language which is very close to embedded C language and it is done on Arduino development environment.

3.3.2 Sensor

Sensor networks are new instruments useful to detect the conditions in remote places in the physical world in environmental monitoring applications such as pollution monitoring, transportation management, intrusion detection and many more [10]. With the constant development in electronics industry, it is possible to collect data remotely and collected data can be transferred to the required platform at a short period of time. There are different sensors available in market which can measure the pollutants and display the value, but the idea here is to select the one which is of low cost and also gives the most accurate values.

Having said that there is a variety of options available for sensors based on the way it measures the pollutants. One such category is Metal Oxide Semiconductor (MOS) gas sensor also known as semiconductor gas sensor, which is used to detect the concentration of any hazardous gases in the atmosphere by changing its resistance. The most popular series available in market for this category is MQ-XX sensors which is popular for its wide

detecting scope, long life, stability, high sensitivity, fast response and also simple drive circuit [1].The sensing material is made up either from Aluminum Oxide (Al_2O_3) or Tungsten trioxide ($WO3$)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 through Platinum wires which is connected to leads made up of Nickel-Chromium, well known conductive alloy. The gas to be detected has a specific temperature at which it gets ionized and the task of the sensor is to work at that temperature. Once the gas gets ionized it gets absorbed by the sensing material which changes the resistance and inturn changes the voltage across the sensor and can be read by the microcontroller [11]. The voltage value along with reference voltage and other resistor's resistance is used to find the resistance of sensor. Once the resistance of the sensor is known then by using the sensitivity curve the concentration could be found out.

Another popularly used MOS sensor which we explored was MICS which are MEMS based whose mode of operation is similar to the above said sensor as both of them are metal oxide. Here, oxidizing gas or the pollutant gas add to the insulative oxygen species causing the resistance to increase [12]. Other than MOS sensors, we also took a look at optical sensor which are spectroscopic devices which uses light scattering principle to find the concentration of pollutants. These sensors are known for its detecting capability of particulate matter of different sizes and is one of the recent development in the field of air quality monitoring. Highly responsive, reliable and long life are the main highlights of this sensor.

Selected Sensor

After understanding the wide range of sensor options in the market the sensors were selected based on their performance, availability, ease of integration, and cost. The selected sensors were:

1. MQ-2 Sensor: This is a a semiconductor gas sensor which has an electrochemical sensor which detects multiple gases such as Carbon monoxide, LPG, methane, and other 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 sensitivity curve for the sensor is shown for different types of gases.

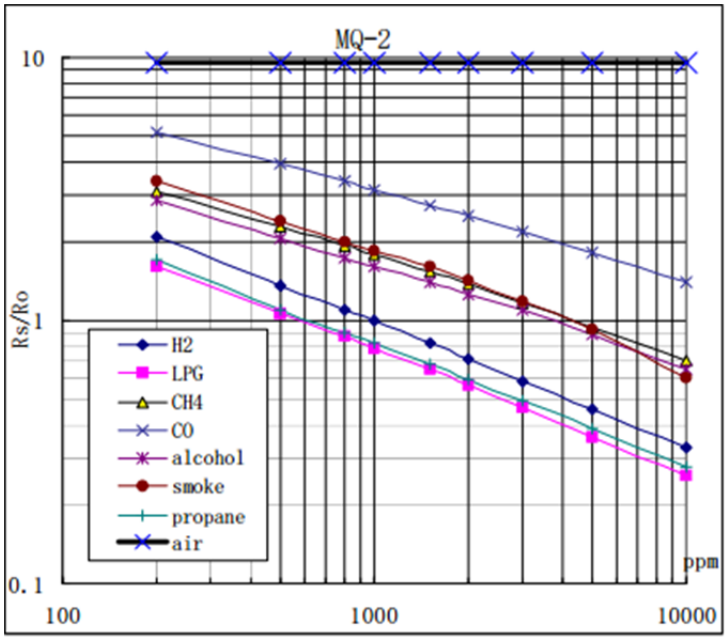


Figure 3.3: Sensitivity characteristics curve [1]

where R_o is the sensor resistance in clean air and R_s is the sensor resistance when exposed to gas.

From the curve in Fig.3.3, the voltage across the sensor is found out depending on which gas one wants to detect and there after using this voltage value the concentration of pollutant is calculated. The range of detection of gas is from 100ppm to 10,000 ppm and has a high sensitivity and fast response time. The sensor is small and portable and provides integration in famous MCU platforms like arduino, raspberry pi etc.

2. MQ 131: Another MQ-XX series sensor that we used for the system is for ozone. The working 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 scope is from 10ppb to 2 ppm of ozone and also has a fast response time and long life.

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.

4. PPD42NS Particulate sensor: The sensor detects the particulate matter through light scattering mechanism and consists of infrared LED positioned in forward angle to a photo diode. As soon as there is a variation in light density, the photo diode detects this and changes the current from the diode [13]. The circuit generates a measurable signal known as Low Pulse Occupancy (LPO) which is proportional to PM concentration [14]. This sensor can measure both PM2.5 and PM10 concentration.

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

3.3.3 Communication Middleware

The collected data needs to be transferred to a platform so that user can understand and interpret the data from the sensor. We have used a WiFi module for this purpose, which will collect the data from the processor and transfers it to a IOT platform said to. The WiFi module used here is ESP8266, which is a highly integrated SOC that meets the requirement of user demand of efficient power usage, compact design and reliable performance [15]. The ESP module can be connected to a processor or it can also be programmed on its own as the module itself is a MCU unit. Unlike the other sensor modules connected to the processor which needs a 5V power supply for its working, this module needs a 3.3V for its power up. The ESP module comes along with installed firmware from AI-thinker and it can be communicated with AT commands.On typing the AT command in the serial monitor the output would come as 'OK' if there is a successful connection.

This firmware can be replaced with user's own code which gives the power of flexibility of connecting with any IOT platform. The arduino platform supports the programming of ESP module which makes the integration much easier. The code can be written in the Arduino IDE and can be flashed into the ESP board by connecting it separately. The circuit for the ESP flashing is as shown which is a voltage divider circuit.[FIGURE NEEDS TO BE INSERTED] When the TX and RX pin of arduino is connected to TX and RX pin of the WiFi module it becomes in the programming mode and flashing occurs.

3.4 System Overview

The complete circuit diagram of how the components are integrated to the Microcontroller is as shown is the figure 3.4. Each of these sensors were carefully arranged so that it reduces the spatial interferences. As majority of the sensors are heat sensors and discharges heat it affects the working of the sensor placed near to it. The initial arrangement of the sensors were done in the breadboard with wires. Building the circuit in breadboard gave the flexibility of working with different arrangements.

After building the circuit in breadboard, it was migrated to a Printed Circuit Board (PCB) developed by Flemings Solution Ltd, India. The design of the board is as shown in the figure 3.5. The PCB was designed in such a way that each sensors and the microcontroller itself was plug-in model to board. This was done so that in case of any malfunctioning of any of the sensors or processor, it 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 zero error circuit and there wont be any short circuiting.

3.5 IoT Architecture

The hardware is the one responsible for the data collection but making these data available to user is done by the visualizing tool. The main objective which was kept when developing

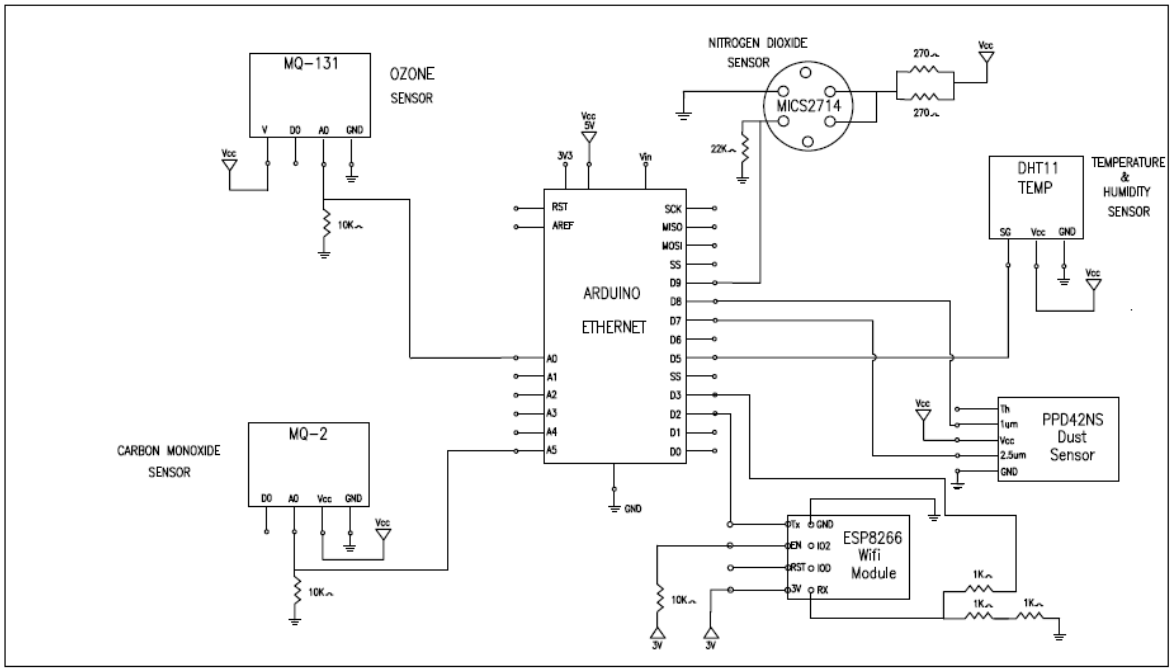


Figure 3.4: System Overview

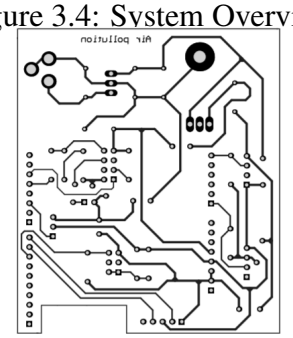


Figure 3.5: Printed Circuit Board

the system was the data collected should be accessed from anywhere and thats why IoT platforms were selected.

Chapter 4

Calibration Of Low Cost Sensor

With the development of sensor technology for air pollution it has attracted a majority of researchers as well as common people to explore and understand more about the pollutants and its effects. This has given freedom for one to set up their own monitoring system at residences, office or even at schools. The problem with this is to identify how accurate the data collected from these commodity sensors to the reference monitoring system. The reference system are the pollution monitoring devices that are developed to a clearly defined standard for a specific criteria pollutant [s] and has completed a rigorous testing and analysis protocol [16]. If the system is giving values which is way too different from the reference value then it brings down the advantages of this technology. This issue can be dealt through calibration which will reduce the uncertainty in data and makes the output more accurate.

Calibration can be defined as the act of evaluating and adjusting the precision and accuracy of measurement equipment [17]. If the measured output from the sensor is not equals to the actual output then it shows that there is a need for calibration. Usually all the electronic instruments are calibrated according to a particular conditions in the laboratory and acquires a certification of calibration before its sold out in the markets. Even then the measured value does not reach accuracy as the condition or the environment where it was calibrated changes that leaves the user with some raw values that gives no information. This issue was taken up and explored by a few researchers in the US Environmental Protection Agency(EPA)

and suggested with three 'Straw-Man Approach' to improve the usability of such data and presented in the Air sensor Workshop [18].

The first approach was by a signal-based calibration technique which requires the data from the remote stations, which is the reference station, to be broadcasted to the local station where the sensor is located and will receive this data and performs a single point calibration of the response. This approach would have been easy if the sensor was already equipped with the data collection and would process automatic calibration.

The next option for calibrating is called the direct sensor calibration that involves placing the sensor in a chamber in which a known concentration of pollutant is set and response is observed. As the concentration of the pollutant is already known the output curve can be compared with it and calibrated accordingly. This is the most common method used for calibration as is often called as laboratory evaluation. Another way of approaching direct calibration is by inspecting the pre-defined response given by the manufacturer and checking how accurate the sensor is to the given concentration. In either case the calibration requires equipment and skills to give accurate concentration value.

The last option is by secondary data normalization in which the concentration values of the pollutant from the low cost sensor is 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 could 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 found out from this. The drawback of this is that the sensor data does not always give a linear response and thus not applicable for all the curves.

Even though these 'Straw-Man Approach' was defined well it was still hard to implement it practically. This eventually led to the development of the 'Air Sensor Toolbox' by the Environmental Protection Agency (EPA) [19] which introduced a tool called 'Macro Analysis Tool' (MAT) for performing comparisons of air sensor data with reference data

and interpreting the results [20].

4.1 Calibration Procedure

The Air Sensor Toolbox provide guidelines for researchers, citizen scientists and developers with working of low cost sensor and its calibration to give more insight about air quality [21]. The Air Sensor guidebook [20] provides a three step procedure for calibrating a sensor:

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

   This means the data collected from the sensor system should be compared with the data from an already calibrated system which is placed by the local authority. This type of comparison is called as 'collocation' and in order to collocate, first find out where the reference system is placed and get access to the data from these system. After getting access to these system place the sensor system at the same level as the standard reference. Once this is done the data collected from both the system can be downloaded in the desired format.

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

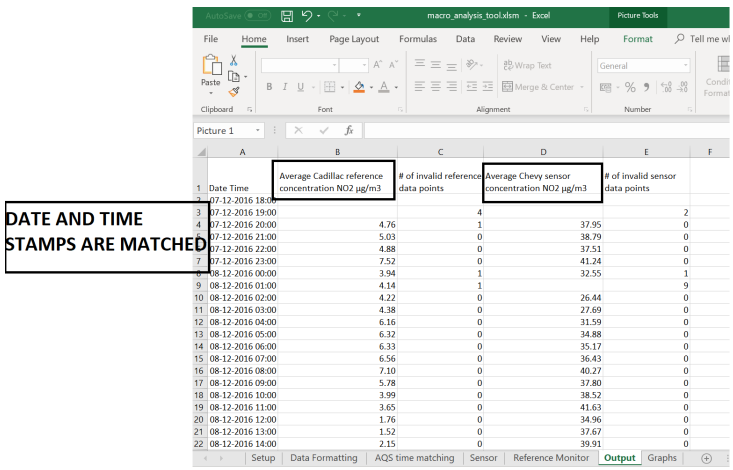
   The relationship between the response of an analytical instrument to the concentration or amount of an analyte introduced into a known instrument is referred to as the "calibration curve" [22]. The MAT tool by EPA is used for drawing the calibration curve and will give an error function as output. This 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 often. There will be changes in the equation and graph each time and this should be noted so as to get accurate value.

4.2 Macro Analysis Tool -MAT

The development of MAT tool gives a huge reliefs for the researchers in calibrating sensor data. This is an Excel based user-friendly macro tool that compares sensor data with reference data [23] even if measurements weren't recorded at precisely the same time, or were collected at different time intervals, such as 1-minute versus 5-minute intervals [24]. The user needs to insert the data from the sensor into the sensor page of the tool and the values from the reference station is inserted into the reference page of the tool. After these values are inserted the details of pollutant, time interval, measurement units and data completeness (amount of usable data obtained) needs to be added in the set up page of the MAT. After all the required fields are filled the output page as shown in the figure 4.1 shows the two sets of data being compared as per the control panel setting made in the tool [23]. This output page basically gives the statistical comparison of two different data sets. From this page it could be seen that the date and time stamps are averaged to a single value for both the system. There is another column in the sheet which is the invalid data points that gives the total number of non validated data points during the observed time. This non validated data points occurs due to either the fault by instrument or it could be the fault in reading by MAT or it could be even due to unacceptable range specified by the user.



	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 4.1: MAT Output

Another output page provided by the tool is correlation graph which plots a graph between the reference monitor and the sensor data as in the figure 4.2. The graph drawn will be a scatter plot and a line called as 'slope-intercept' will be drawn through the data points and is represented simply by the equation $y = mx + c$. This equation represents the average behavior of the sensor data (vertical axis, represented by y) compared with the reference data (horizontal axis, represented by x) [23]. The slope of the graph shows the similarity or difference in sensor measurements when compared with the reference instrument, on average.

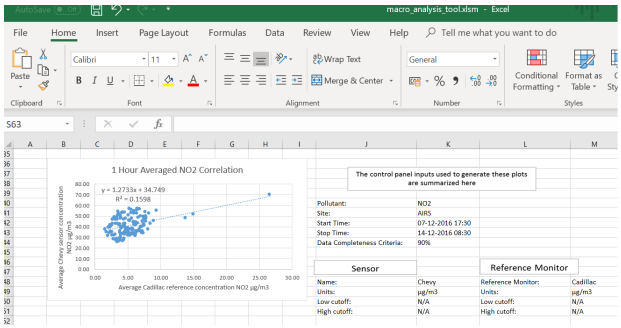


Figure 4.2: Correlation Graph

The tool also generates a coefficient of determination (also known as correlation coefficient in statistics) or R-Squared (R^2) shows how close the value is to the slope-intercept line. The value ranges from 0 to 1 and the closer R^2 is to 1, the stronger the agreement between the sensor and the reference data [25]. Along with these output the tool also generates a time series output graph which will show the concentration of pollutant for both the system as in the figure 4.3

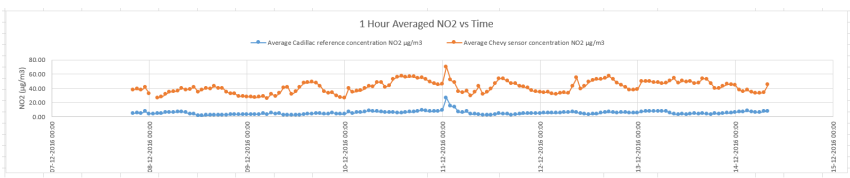


Figure 4.3: Time-Series Graph

From the Time-Series graph it could be seen how the closely value from two different system is related to. Having included all these functionality in the tool gives a strong solution for the calibration. The development of the tool was a year-long project of EPA with

other community group(Clean Air Carolina) and one tribal nation(Eastern Band of Cherokee Indians). The next section will expand in detail about the background concept about the working of tool which is linear regression.

4.3 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) [26]. 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 4.4 could be drawn through it by Least Square method.

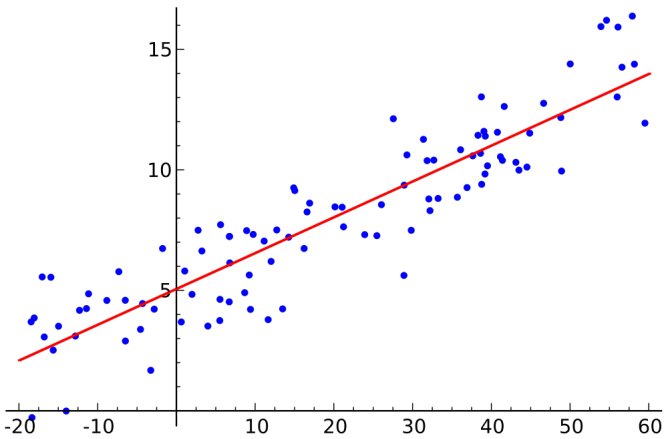


Figure 4.4: Linear Regression Model [2]

The method of Least Squares is a procedure to determine the best fit line to data [27] 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 [28] [29]. The line drawn will have a slope m

and is given by the formula

$$m = \frac{\sum_i[(x_i-\bar{x})(y_i-\bar{y})]}{[\sum_i(x_i-\bar{x})^2]} \text{ [30] [31] [26]}$$

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

$$C = \bar{y} - m\bar{x} \text{ [30] [32]}$$

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 [30]. The equation for finding the regression value is

$$R^2 = \frac{\sum_i[(x_i-\bar{x})(y_i-\bar{y})]}{\sqrt{[\sum_i(x_i-\bar{x})^2[\sum_i(y_i-\bar{y})^2]]}} \text{ [30]}$$

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.

Bibliography

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

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

[3] 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.

[4] Frequently Asked Questions. The air quality health index. *Air Quality health index FAQ*.

[5] J S Asha and P Sindhu. Website : www.ijirset.com Assessment of Air Quality in Two Cities of Kerala based on AQI by USEPA Method and. (2009):9284–9292, 2017.

[6] Debanshee Saha. IoT based Air Quality Monitoring System using Wireless Sensors deployed in Public Bus Services. 1952.

[7] Various Contributors. Essay on Air pollution: Introduction, Causes, Sources, Impact and Control Measures. 2013.

[8] 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.

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

[10] Young Jin Jung, Yang Koo Lee, Dong Gyu Lee, Yongmi Lee, Silvia Nittel, Kate Beard, Kwang Woo Nam, and Keun Ho Ryu. Design of sensor data processing steps in an air pollution monitoring system. *Sensors*, 11(12):11235–11250, 2011.

[11] 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].

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

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

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

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

[16] Eric S Hall, Surender M Kaushik, Robert W Vanderpool, Rachelle M Duvall, Melinda R Beaver, Russell W Long, and Paul A Solomon. Federal Reference Method (FRM), Federal Equivalent Method (FEM), National Ambient Air Quality Standards (NAAQS), Sensors; Federal Reference Method (FRM), Federal Equivalent Method (FEM), National Ambient Air Quality Standards (NAAQS), Sensors. *American Journal of Environmental Engineering*, 4(6):147–154, 2014.

[17] Davis) Kejuruteraan, Fakulti Tekn, NIST Engineering Statistics Handbook Jenny Wu (University of California and 3. Instrument Calibration. pages 1–7, 2018.

[18] Ron Williams, Tim Watkins, and Russell Long. Low Cost Sensor Calibration Options The Straw-Man Approach. 2013.

[19] Ron Williams, Vasu J Kilaru, Emily G Snyder, Amanda Kaufman, Timothy Dye, Andrew Rutter, Ashley Russell, and Hilary Hafner. Air Sensor Guidebook. *Epa/600/R-14/159*, (1):1–5, 2014.

[20] 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].

[21] 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].

[22] Epa and Ord. Calibration Curves: Program Use/Needs Final. (October), 2010.

[23] 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.

[24] 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.

[25] Ron Williams. Emerging Sensor Technologies. 2018.

[26] Shaun Burke and High Wycombe. Regression and Calibration. pages 13–18.

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

[28] 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.

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

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

[31] Diana Arsova, Sofia Babanova, and Petko Mandjukov. Linear Calibration – Is It so Simple ? pages 35–42, 2009.

[32] David Harvey. 5.4: Linear Regression and Calibration Curves - Chemistry LibreTexts. *LibreTexts*, pages 1–12, 2016.