Report for end-semester evaluation of CE499 course

Identification of Hotspot and Spatial Analysis for Particulate Matter and Gaseous Pollutants.

Submitted

By

Gourav Dangi

Under the supervision of

Dr. Sharad Gokhale

Department of Civil Engineering

Indian Institute of Technology Guwahati

April 2023

CERTIFICATE

It is certified that the work contained in the project report entitled "**Identification of hotspot and Spatial Analysis for Particulate Matter and Gaseous Pollutants**", by **Gourav Dangi** (Roll No: 190104035) has been carried out under my/our supervision and that this work has not been submitted elsewhere for the award of a degree or diploma.

Dr Sharad Gokhale

Department of Civil Engineering

Indian Institute of Technology Guwahati

ABSTRACT

The overarching aim of this work is to develop a Predictive Model to identify Air Pollution Hotspots and predict whether the pollutants are exceeding standard permissible limit. Air pollution hotspots were identified earlier using rigorous mathematical methods, but using modern computational methods to aid remains a key topic that is unexplored in this area. Poor air quality threatens the health of all living things from humans, plants to animals. Air pollution is a growing concern for all nations, and more than 7 million people die from air pollution each year. This project will help to identify regions with pollution and the pollutant which is a major concern so appropriate air quality management can take place.

In my study I have chosen Guwahati for finding air quality. Only 6 monitoring stations are available in Kamrup. I have covered 28 km X 10 km area for study purposes including all 6 monitoring areas inside my study region. Air quality has been worsening as time passes in Guwahati.

ACKNOWLEDGEMENT

I would like to thank **Dr. Sharad Gokhale** for guiding me and providing me with some helpful suggestions as well as a lot of encouragement. Working under his direction has been a fantastic experience, and I am grateful for his trust in me and my work. This work would not have been completed in such an easy and straightforward manner if it had not been for his inspiration and assistance. I am thankful to Rajarshi Sharma for his valuable suggestions, guidance and valuable support which helped me in completing my project work.

Date: Signature of the student

Gourav Dangi

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

Air pollution is one of the most serious threats to the environment on a global scale. As per WHO report more than 7 million people die every year due to air pollution. As a sign of modernization, urbanization, up-gradation of lifestyle and increased population and traffic volume the concentration of various pollutants has increased to a point which is beyond safe. The increasing concerns about air pollution and its adverse effects on human health and the environment have led to extensive research in the field of air quality management. In particular, the identification of hotspots and spatial analysis for particulate matter and gaseous pollutants has been a topic of great interest. The results of these studies can be used to develop effective strategies to reduce the emissions of particulate matter and gaseous pollutants and mitigate their adverse effects on human health and the environment.

Air Quality is deteriorating due to increased pollution level, major criteria pollutant includes particulate matter (PM either PM2.5 or PM10), sulphur dioxide (SO2), nitrogen monoxide (NO), Nitrogen Dioxide (NO2), Carbon monoxide (CO) and many more. Monitoring stations measure the pollution level at a particular fixed location but air pollution is occurring over a wide range of areas. The source of emission ate not only stationary such as industry but also mobile such as traffic. As a result, the optimum method is to map using an interpolation technique based on available data.

Computation power has significantly increased in industry and with modern tools and technology we can predict the upcoming result with high accuracy. Therefore, it is very important for us to keep up industry trends and use this advancement in technology to our advantage. The data collected was not standardized and incomplete. Significant data processing was performed whose details are not discussed in brief. Missing values can be replaced with mean without any problem as the curve is smooth. CO2, PM2.5 were not used as a pollution parameter as their data was mostly missing from the monitoring stations.

High levels of Pollutants and associated health risk are major concerns for the government. Due to its toxic nature PM2.5 has been reported as a crucial pollutant in urban areas, it alone has caused 0.67 million deaths in India in 2017. The city management and people in-charge should be aware of the hotspot regions and the specific pollutants which are exceeding the limit so appropriate and effective measures can be taken.

Chapter 2

Literature Review

In-situ measurements, air quality models, and satellite observation are all possible sources of air quality data. However, none of these data sources are acceptable for mapping purposes on their own, due to large data gaps, insufficient geographic resolution, or major uncertainties. Several studies have been conducted to identify hotspots for particulate matter and gaseous pollutants.

As per Goyal et al. (2021) Critical assessment of spatio-temporal variations in pollution levels is a crucial step for identifying and prioritizing air pollution hotspots (APH) in urban areas. That study concluded the need for proper management especially during the month of October – February where additional efforts are needed to combat the concerning pollution.

V. Todorov, I. Dimov conducted a comprehensive experimental study of Monte Carlo algorithm based on modifications of the Latin Hypercube Sampling Edge and Random Algorithms for multidimensional numerical integration and Latin hypercube was better in fewer iterations for large data.
Latin Hypercube Sampling is typically used to save computer processing time when running Monte Carlo simulations. Studies have shown that a well-performed LHS can cut down on processing time by up to 50 percent (versus a standard Monte Carlo importance sampling) Xin, L. (2014).

Similarly, a study conducted by Li et al. (2020) used a spatial clustering method to identify hotspots of PM2.5 in the Yangtze River Delta region of China. The study identified several hotspots of PM2.5 concentrations, with the highest concentrations found in the industrial areas.

A study conducted by Yang et al. (2020) used a hybrid approach combining spatial clustering and regression modelling to identify hotspots of PM2.5 in the Beijing-Tianjin-Hebei region of China. The study identified several hotspots of PM2.5 concentrations, with the highest concentrations found in the urban areas.

Traffic vehicular emissions are one of the major pollutant emission sources. In western Europe automobiles have surpassed industrial activities and coal burning is the primary source of air pollution (Namdeo et al., 2002). According to Ramachandra et al. (2009) in India, passenger cars (PCs) emit roughly 34% and 50% of CO and NO_x, respectively.

Dubey et al., (2021) studied that those emissions during special occasions may vary, so prediction cannot be validated. Terrain data are not included for air quality mapping, though the air pollution prediction model along with interpolation increases the reliability of the report.

2.1 Air Pollution Hotspots

Air pollution hotspot refers to a particular area or location where the concentration of pollutants in the air is significantly higher than the surrounding areas. These hotspots can be caused by various sources of pollution, including industrial emissions, traffic, and natural phenomena such as wildfires.

Hotspots can be detrimental to human health and the environment as they can cause respiratory problems, cardiovascular disease, and other health issues. The pollutants can also harm vegetation and contribute to climate change.

Identifying air pollution hotspots is crucial in developing effective strategies to reduce pollution levels and protect public health. Governments and organizations can use various methods to identify hotspots, including air quality monitoring, satellite imagery, and modeling techniques.

Reducing air pollution in hotspots requires a multi-faceted approach that involves regulating emissions from sources such as factories and vehicles, promoting cleaner energy alternatives, and educating the public on the impacts of air pollution. By taking action to reduce pollution in hotspots, we can improve the health and well-being of communities and protect the environment.

2.2 Kernel Density Estimation

Kernel density estimation is a statistical method used to estimate the probability density function of a continuous random variable. In simple terms, it is a way to estimate the shape of a probability distribution based on a set of observed data points.

The choice of kernel function can affect the accuracy of the estimation, but popular choices include the Gaussian and Epanechnikov kernels. The width of the kernel, known as the bandwidth, also affects the accuracy of the estimation, with smaller bandwidths resulting in a more detailed estimate but also more noise.

Kernel density estimation has a wide range of applications in various fields, including ecology, finance, and image processing. It can be used to analyze and visualize data, identify patterns and trends, and make predictions about future events.The method requires calculating the kernel function for each data point, and the number of calculations can quickly become unmanageable as the dataset grows. Additionally, choosing an appropriate bandwidth for the kernel can be challenging, as a bandwidth that is too small can result in overfitting and a bandwidth that is too large can result in underfitting. This can lead to biased estimates of the probability distribution and reduce the accuracy of the method.KDE assumes that the underlying distribution is smooth and continuous, which may not always be the case in real-world data. In cases where the distribution has a significant amount of noise or irregularities, alternative methods such as histogram or binning may be more appropriate.

Due to the complex nature and high computation power using KDE was avoided but It remains a useful tool with respect to study of work.

Chapter 3

Methodology

3.1 Study area

Kamrup Metropolitan district is one of 36 districts of Assam and has a population of over 10 lakhs, according to the Indian census 2011. The population density of the district is 820 people per square km and it covers an area of 1527.84 square km. Guwahati is the most important city in north eastern India in every aspect. It has a population of over 9 lakhs according to the Indian census 2011. Guwahati is administrative capital of Kamrup (M) district.

The study area covers a total of 28 km long and 12 km wide rectangular area extending between north latitude of 26° 06’ 13.65’’ and 26° 12’ 44.62’’ and east longitude 91° 38’ 28.48’’ and 91° 50’ 40.59’’ forming a rectangular region. It includes a total of 6 monitoring stations of PCBA out of which 3 are continuous ambient air quality monitoring stations.

These stations are as follows:

1. PCBA Head Office Bamunimaidan
2. Industrial Training Institute Gopinath Nagar
3. Guwahati University

Three continuous ambient air quality monitoring station (CAAMS) are:

1. IIT Guwahati
2. Pan Bazaar
3. Railway colony

The study region is divided into 14 * 5 grids of 2km x 2km each using ArcMap 10.8

Fig 1: study region with air monitoring station

3.2 Data Collection and Processing

The air quality data was collected from the CPCB website. The three significant pollutants are recognized as NO2, SO2, PM including PM2.5 and PM10. The Concentration data of air pollutants namely NO2, SO2, PM2.5 and PM10 was collected for a duration of 1 year (January 1st 2022– December 31st 2022).

The Collected data was not usable and had to be processed thoroughly and converted into usable form. Emission data was in “25 file” format which was converted into text and then csv, which was then used to find results. The data itself had flaws apart and had to be manipulated (within acceptable limits) to use and gain results. Lack of data was a major issue and because of which PM2.5, CO, NO were dropped altogether.

3.3 Data Analysis and Interpretation

The criteria air pollutants SO2, NO2 and PM (both PM2.5 and PM10) were taken from CPCB and PCBA websites. The data collected for the duration of 1-year Jan 1 2022 to Dec 31 2022.

For Guwahati we tried to find trends of these pollutants around the year and how seasons affect the pollutants. From the finding of previous data interpretation, we found that concentration of PM10 exceeds their permissible limit of 60µg/m3 often whereas SO2 and NO2 were mostly under their limits of 50µg/m3 and 40µg/m3, however during winters their concentration also exceeds some places.

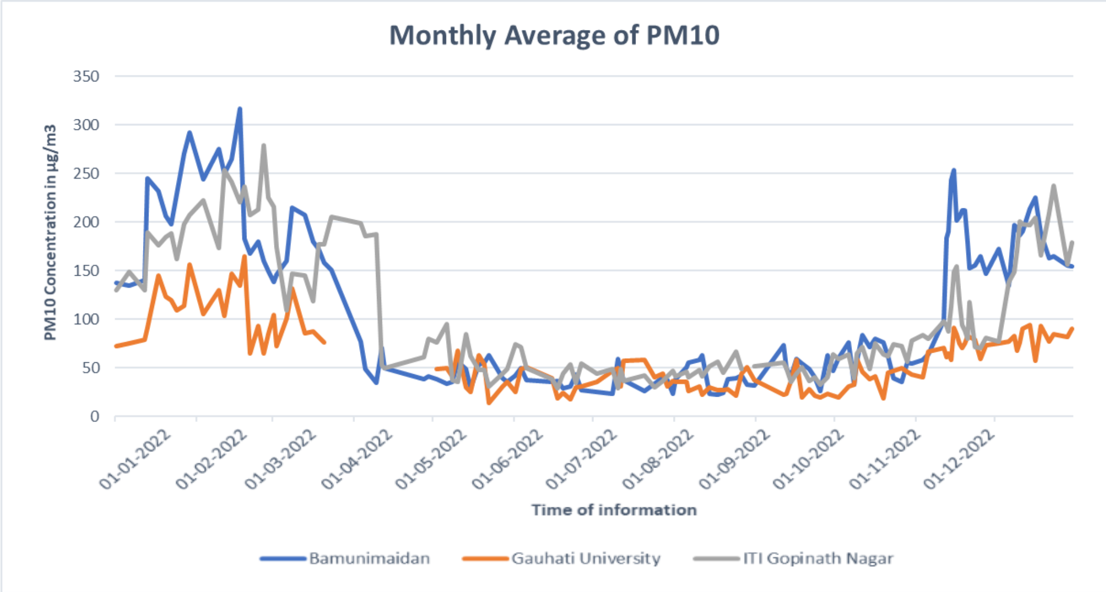

Fig 2:Monthly Average PM10 between Bamunimaidan, Gauhati University and Gopinath Nagar

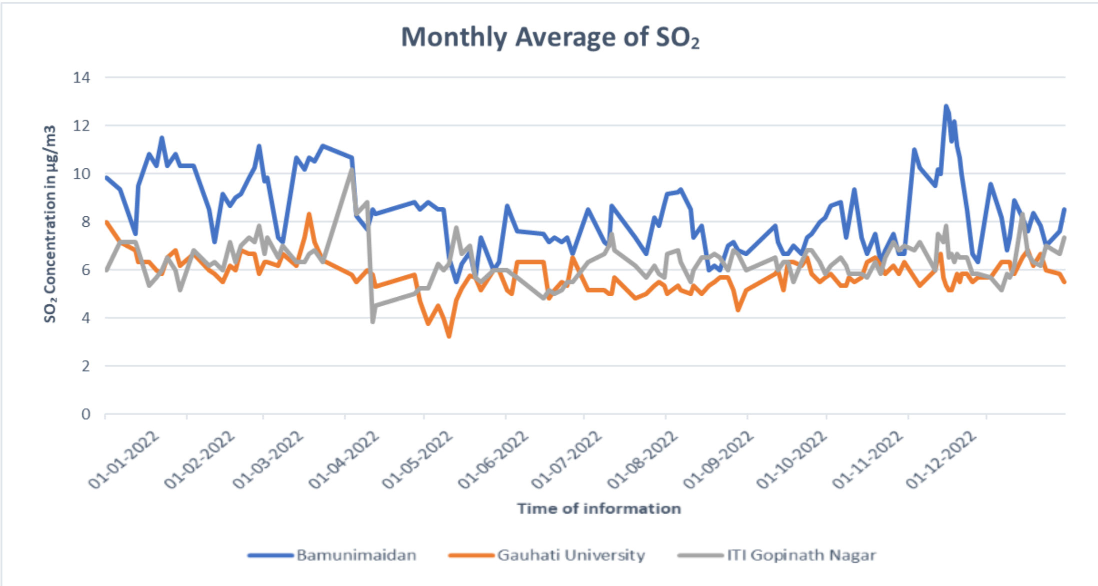

Fig 3: Monthly Average SO2 between Bamunimaidan, Gauhati University and Gopinath Nagar

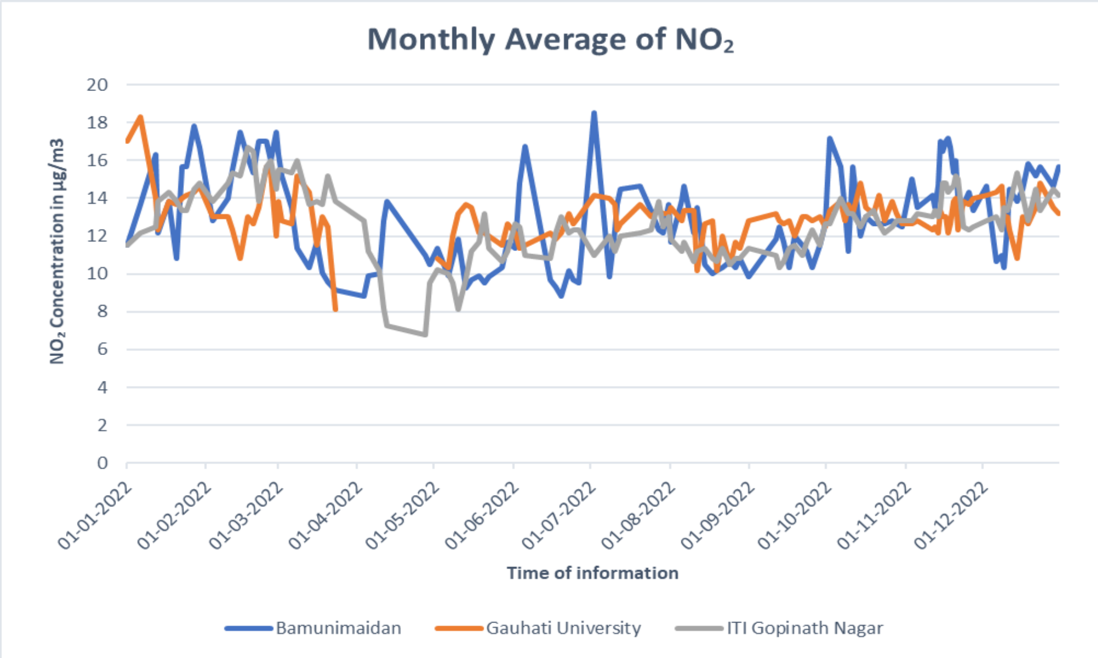

Fig 4: Monthly Average NO2 between Bamunimaidan, Gauhati University and Gopinath Nagar

A rise was observed during winters (post mid-November to early mid-march) which was followed by a steep decline.The main cause of NO2 in air pollution is combustion of fossil fuel simply put in traffic. Such seasonal variation can also be found in residential zones but the levels are mostly stable unlike here. SO2 follows a trend similar to that of PM with rise during monsoon-winters followed by a decline during summers.

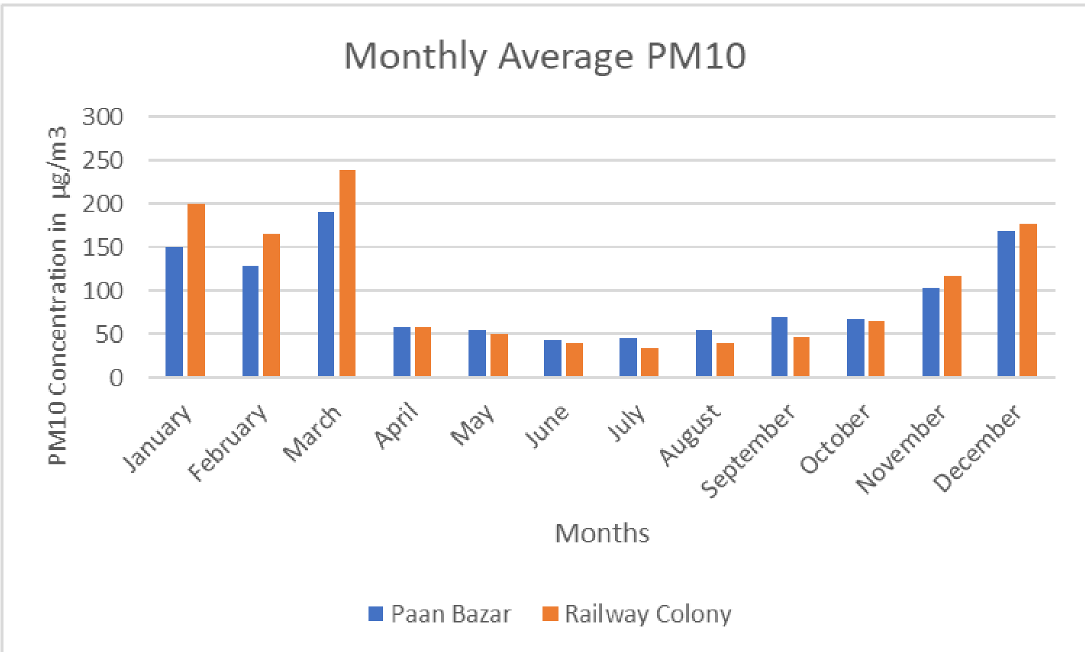

Fig 5:Monthly Average PM10 between Pan Bazaar and Railway colony

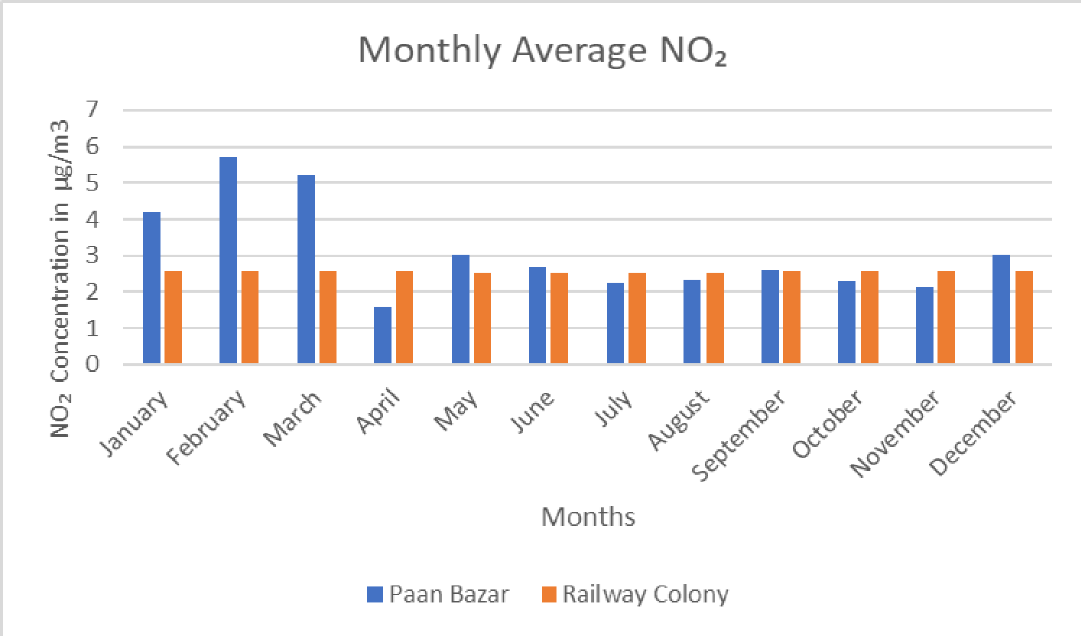

Fig 6:Monthly Average NO2 between Pan Bazaar and Railway colony

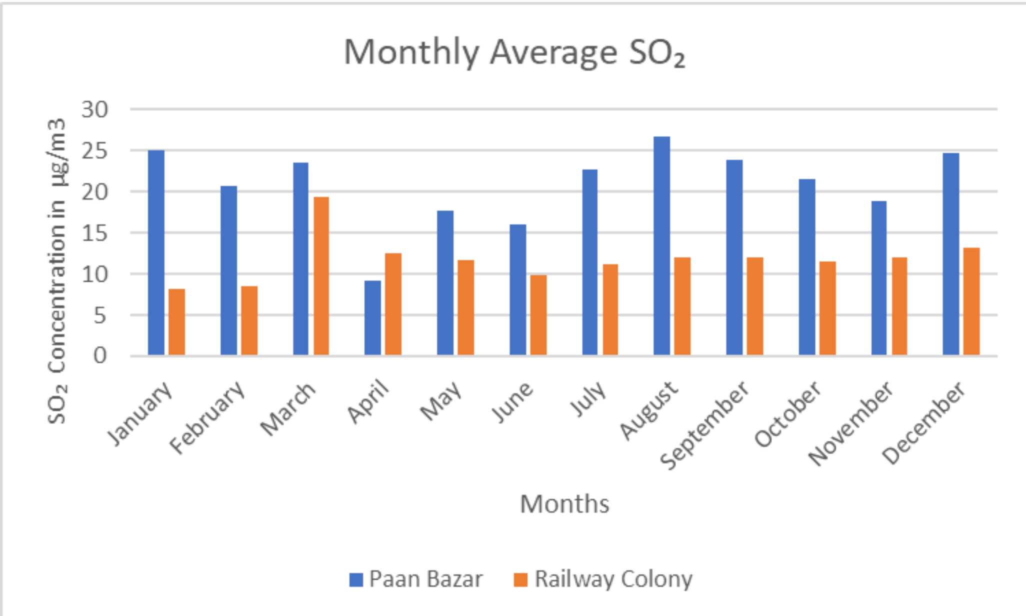

Fig 7:Monthly Average SO2 between Pan Bazaar and Railway colony

A similar trend was observed for Pan Bazaar and Railway colony. As they were CAAMS stations and had complete data available their plots were created separately.

The concentration is well within the standard for summers. The variations observed were due to synergistic effects of localized sources, festivals, and unfavorable meteorological conditions. The AQI was high during these periods. Thus it can be inferred the data for monsoon-winters is suited for preventive measures as they represent the maxima of data.

3.4 Value Prediction and Heat map Generation

Concentration and Emission are directly proportional so instead of pure random sampling a latin hypercube type model was used to interpolate/find missing data.

Emission data of the entire working region was known and concentration data was known for only 6 blocks of grid, which were the six monitoring stations Randomly known data points were chosen, and as concentration and emission are proportionate, missing emission was found using a BFS (breadth first search) run. This process was repeated 6 times and an average value was taken to predict the final emission value of the block and eventually the Grid.

Fig 8: region with known concentration value

Procedure to generate heat map:

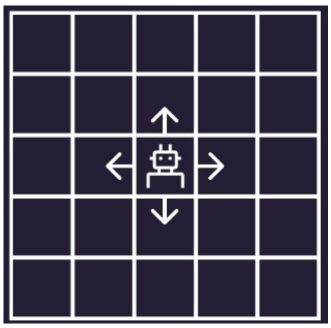

Fig 8.1: BFS representation how data move

1. The concentration was known sparsely as mentioned in fig 8 however emission data was known for all blocks of grid.
2. Randomly 1 out of 6 monitoring stations (block in grid) was chosen and the BFS (breadth first search) model was run. This method closely resembles the Latin hypercube and random points were chosen.
3. Concentration and Emission are proportionate, so c1/e1 = c2/e2. Using this relation all unknown blocks of grids are predicted.

4. Starting points do affect the predicted value a little hence this operation was run 6 times (one for each monitoring station) and average values were taken.
5. Heat map was plotted after prediction of data and hotspots (The region in red) were identified.

Results and Discussion

AQI is an index for reporting air quality daily. It measures how air pollution affects one's health within a short period of time. Its purpose is to help people know how the local air quality impacts their health. Considering only the pollutants in our study

$$AQI = \frac{1}{3}\left(\frac{(NO2)obs}{(NO2)std} + \frac{(SO2)obs}{(SO2)std} + \frac{(PM10)obs}{(PM10)std}\right) * 100 \quad \text{(i)}$$

where obs prescript denotes observed value and std represent standard permissible limit by NAAQS for each pollutant respectively.

The air quality of Guwahati and its nearby area comes under either good (0-50) or moderate (100-200) range most of the time, however air quality in this study area is poor (200-300) or very poor (300-400). The concentration of pollutants exceeds their permissible limit or stays close to it. The study area includes mainly residential and commercial areas. The levels of PM10 and PM2.5 were very high especially at Railway Colony and Pan Bazaar.

SO_2 levels were under limits and were in range 8-12 μg/m3 which was under the permissible limit of 50 μg/m3. NO_2 levels were under limits and were in range 13-24 μg/m3 which was under the permissible limit of 40 μg/m3.

The transition period of changing air quality is October and April month and the pollutant concentration was significantly lower during April - October period.

Railway colony and southern region of our study area namely Gopinath Nagar and Bamunimaidam were identified as hotspot, The area identified as hotspot has a nearby industrial cluster with various factories like IOCL Refinery, Indian Carbon Factory, Assam Carbon Limited and Assam Roofing Pvt Ltd etc. which can be the reason for high pollutant levels.Other meteorological parameters such as temperature, wind speed and wind direction may also have influenced the concentration of pollutants to a significant level which are not included in the study.

Resulting Plots for each month of 2022:

all measurements in µg/m³ unless mentioned otherwise

JAN 2022

PM10

Fig 9: heat map for PM10

SO2

Fig 10: heat map for SO2

NO2

Fig 11: heat map for NO2

Hot spots were identified at Railway Colony and surrounding region for PM10 and SO2, and at Gopinath Nagar for NO2. NO2 primarily gets in the air from the burning of fuel. NO2 forms from emissions from cars, trucks and buses, power plants, and off-road equipment, whereas PM10 is caused by industry, dust etc.

FEB 2022

PM10

Fig 12: heat map for PM10

SO2

Fig 13: heat map for SO2

NO2

Fig 14: heat map for NO2

Hot spots were identified at Railway Colony and Gopinath Nagar for PM10, for NO2 hotspots formed at Gopinath Nagar, Railway Colony and outer edge of IIT Guwahati. SO2 hotspots were confined to Pan Bazaar and Railway Colony.

MAR 2022

PM10

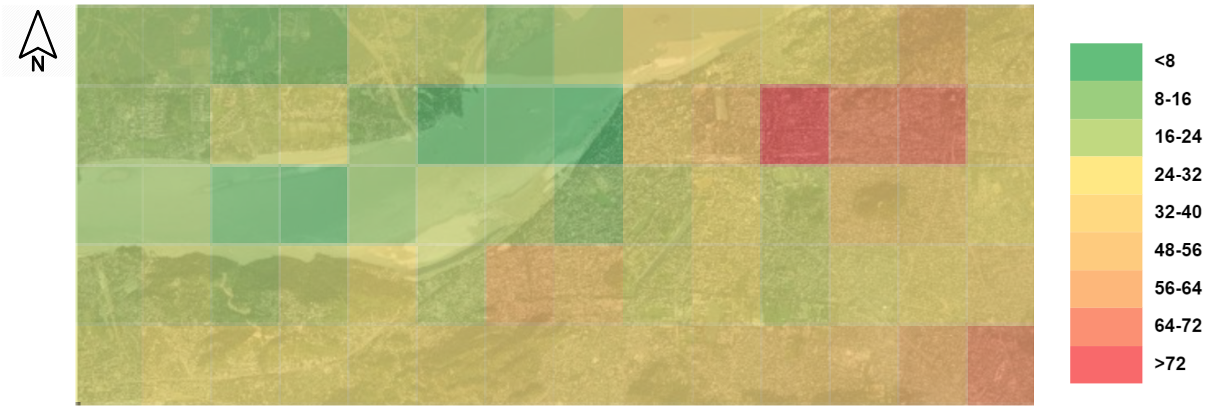

Fig 15: heat map for PM10

SO2

Fig 16: heat map for SO2

NO2

Fig 17: heat map for NO2

For PM10 Hotspots were identified at Railway Colony and surrounding regions, for NO2 hotspots were allocated at Bamunimaidan and Gopinath Nagar, For SO2 a number of places had higher concentrations, the most notable hotspot were at GopinathNagar and surrounding regions.

APR 2022

PM10

Fig 18: heat map for PM10

SO2

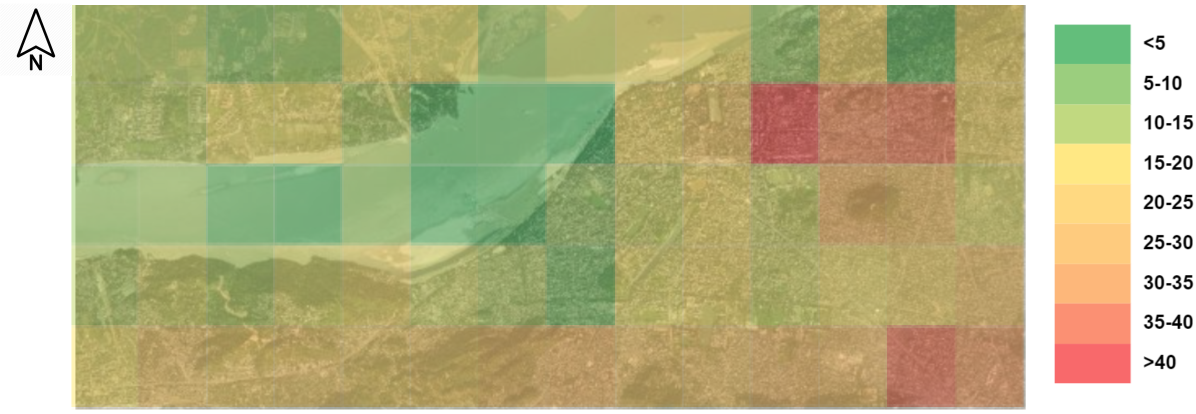

Fig 19: heat map for SO2

NO2

Fig 20: heat map for NO2

For PM10 Hotspots were identified at Railway Colony and Gopinath Nagar, for NO2 hotspots were allocated at Gopinath Nagar, For SO2, the most notable hotspots were at Railway Colony, a minor hotspot was also noted at Bamunimaidan.

MAY 2022

PM10

Fig 21: heat map for PM10

SO2

Fig 22: heat map for SO2

NO2

Fig 23: heat map for NO2

For PM10 multiple places had higher concentration, hotspots were identified at Railway Colony, for NO2 hotspots were allocated at Gopinath Nagar, For SO2, the most notable hotspots were at GopinathNagar and surrounding regions.

June 2022

PM10

Fig 24: heat map for PM10

SO2

Fig 25: heat map for SO2

NO2

Fig 26: heat map for NO2

For PM10 multiple places had higher concentration, hotspots were identified at Railway Colony and surrounding regions, for NO2 hotspots were allocated at Gopinath Nagar, For SO2 a number of places had higher concentrations, the most notable hotspots were at Bamunimaidan and GopinathNagar and surrounding regions.

July 2022

PM10

Fig 27: heat map for PM10

SO2

Fig 28: heat map for SO2

NO2

Fig 29: heat map for NO2

For PM10 Hotspots were identified at Railway Colony and surrounding regions, for NO2 hotspots were allocated at Gopinath Nagar, For SO2 hotspots were at GopinathNagar and surrounding regions.

August

PM10

Fig 30: heat map for PM10

SO2

Fig 31: heat map for SO2

NO2

Fig 32: heat map for NO2

For PM10 multiple places had higher concentration, hotspots were identified at GopinathNagar, Guwahati University, IIT Guwahati also had relatively higher concentration, for NO2 hotspots were allocated at Gopinath Nagar, For SO2 hotspots were at Paan bazar and Railway Colony.

September 2022

PM10

Fig 33: heat map for PM10

SO2

Fig 34: heat map for SO2

NO2

Fig 35: heat map for NO2

For PM10 multiple places had higher concentration, hotspots were identified at Railway Colony, for NO2 hotspots were allocated at Gopinath Nagar, For SO2, the most notable hotspots were GopinathNagar.

October 2022

PM10

Fig 36: heat map for PM10

SO2

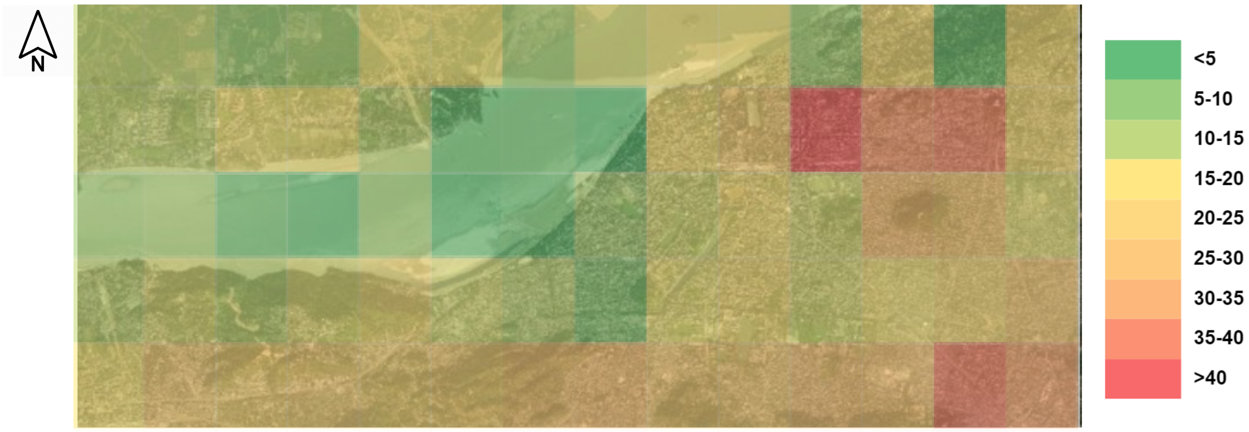

Fig 37: heat map for SO2

NO2

Fig 38: heat map for NO2

For PM10 multiple places had higher concentration, hotspots cannot be pinpointed, for NO2 hotspots were allocated at Gopinath Nagar, For SO2 hotspots were at Bamunimaidan and Railway Colony.

November 2022

PM10

Fig 39: heat map for PM10

SO2

Fig 40: heat map for SO2

NO2

Fig 41: heat map for NO2

For PM10 hotspots were identified at Railway Colony and a notable rise in concentration was observed at IIT Guwahati for NO2 and So2 hotspots were allocated at Gopinath Nagar.

December 2022

PM10

Fig 42: heat map for PM10

SO2

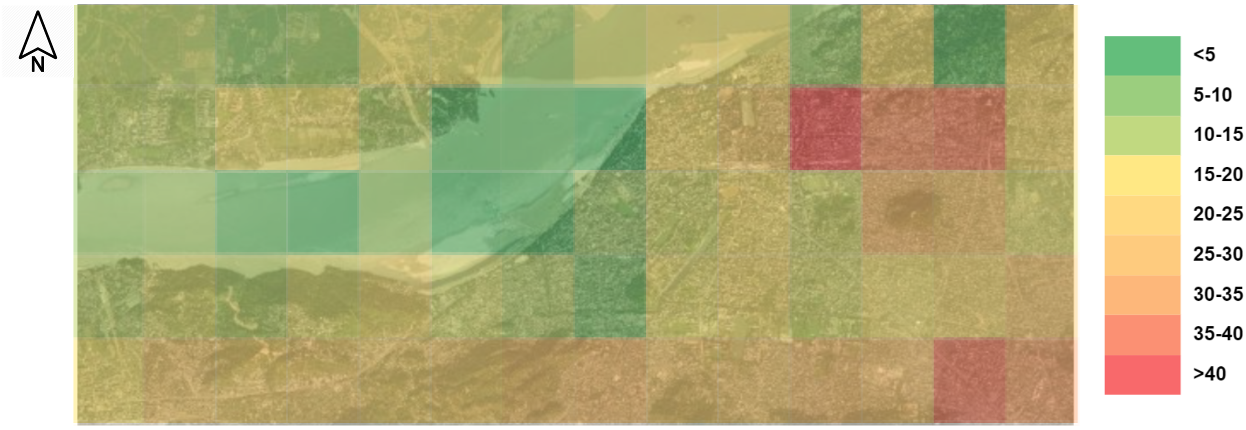

Fig 43: heat map for SO2

NO2

Fig 44: heat map for NO2

For PM10 multiple places had higher concentration, hotspots were difficult to pinpoint, for NO2 hotspots were located at Gopinath Nagar, For SO2 hotspots were noted at Railway colony.

Conclusion and Future Scope of Work

The months of October and April are transitional for changing air quality, and the concentration of pollutants was noticeably lower from April to October.
The railway colony and southern part of our study area, Gopinath Nagar and Bamunimaidam, were designated as hotspots. The area designated as a hotspot has an industrial cluster nearby with a number of factories, including the IOCL Refinery, Indian Carbon Factory, Assam Carbon Limited, and Assam Roofing Pvt Ltd, which may be the cause of the high pollutant levels. Although they were not considered in the study, additional meteorological factors like temperature, wind speed, and wind direction may also have had a significant impact on the concentration of pollutants.

The concentration values were found and plotted, and in some regions values of PM10 and NO_2 were exceeding the standard permissible limits. These regions are classified as hotspots.In our study we found that Railway colony is a persistent hotspot and its air quality comes under poor - very poor category. Rest of the study region falls under good, moderate and satisfactory. During Monsoon air quality falls under the good category so it is not a matter of concern.

Concentration and emission data can be collected on a large scale to find the values more accurately and for smaller grid size.

Mathematical models can be implemented to cross validate our prediction like IDW (inverse distance weighting) or Kriging and use of techniques like KDE can improve predictions.

Since 2 more new CAAQMS stations have been installed (IIT Guwahati and LGBI airport) The quality of data is improving and after a significant amount of data is collected it can be incorporated in our study.

Grids can be refined and made smaller with proper emission data to carry further analysis.

Regions can be divided and studied individually, such as residential (Railway colony), Institutional (IIT Guwahati, GU, ITI), Comercial/Tourist (Pan Bazaar).

References

Alexeeff, S.E., Schwarts,J., Kloog,I.,Chudnovsky, A.koutrakis,P., & Coull, B.A. (2015). Consequences of kriging and land use regression for PM2.5 prediction in epidemiologic analyses: Insight into spatial viability using high resolution satellite data. Journal of exposure Science and Environmental Epidemiology 25(2),138-144

Lin, Y.-P.; Chu, H.-J.; Wu, C.-F.; Chang, T.-K.; Chen, C.-Y. Hotspot Analysis of Spatial Environmental Pollutants Using Kernel Density Estimation and Geostatistical Techniques. Int. J. Environ. Res. Public Health 2011, 8, 75-88

Prachi Goyal, Sunil Gulia, S.K. Goyal, "Identification of air pollution hotspots in urban areas - An innovative approach using monitored concentrations data ", ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2021.149143

Moldovan, Clementina & Sbirna, Sebastian & Sbirna, Liana-Simona. (2022). Presenting air quality status through hot spot maps realized by using Kernel Density Estimation (KDE) - case study: Craiova, 2020/2021. MATEC Web of Conferences. 373. 00067. 10.1051/matecconf/202237300067.

Awkash Kumar and Indrani Gupta and Jørgen Brandt and Rakesh Kumar and Anil Kumar Dikshit and Rashmi S. Patil, Air quality mapping using GIS and economic evaluation of health impact for Mumbai City, India, Journal of the Air & Waste Management Association 66(5), 470-481

M,S., & Q,A.(2015), Application of GIS in Visualization and Assessment of ambient Air Quality for SO_2 and NOx in Sheikhpura city, Pakistan, journal of Geography and natural disaster, 05(03)