Stationary vs Mobile IoT Nodes: A Spatio-Temporal Analysis of PM, Air Pollution Data

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Electronics and Communication Engineering
by Research

by

Souradeep Deb
2019702004
souradeep.deb@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
November 2022

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Stationary vs Mobile IoT Nodes: A Spatio-Temporal Analysis of PM, Air Pollution Data" by Souradeep Deb, has been carried out under our supervision and is not submitted elsewhere for a degree.

Date	Advisor: Dr. Sachin Chaudhari
Date	Co-Advisor: Dr. K.S. Rajan

To

My Family and Friends

Acknowledgments

As I reminisce the day I cracked the PGEE exam and came to IIIT Hyderabad in August 2019 to pursue the MS by research program, I still feel the same sense of exhilaration and excitement. As I was entering the campus, I realized I've finally embarked on a journey of learning and exploring cutting edge technology that is going to be life changing. As I am on the onset of my third year at IIIT-H, I feel my overall experience has been very constructive, fruitful and full of exciting adventures. Although there have been many hurdles, but I will forever be indebted to fate for providing me the opportunity to be here and make the most out of it.

Firstly, I would like to express my deepest gratitude to my research advisor, Dr. Sachin Chaudhari. During my first semester, when I approached to work under him as a research assistant, I was a novice with little to no experience about research. Not only did he immensely supported me during the early stages, but his holistic guidance and hands-on approach has led me to complete my research with ease. I shall forever be grateful to him for the experience I garnered working in his research group. His relentless pursuit and sheer passion for his work has inculcated in me a hunger for solving real life problems through technology-based solutions. I want to thank him for believing in my skills, and enabling me to evolve into the researcher and person that I am today.

Secondly, I would like to thank my co-advisor, Prof. K.S. Rajan for his support and guidance throughout my research. Just like my advisor, he also guided me since my early days of research by providing his domain expertise that enabled me to incorporate innovative ideas and methods in my research. I owe it to him for broadening my understanding of problem viewing and analysis. The lessons that he imparted during our discussions on systematic approach of dealing with the problem in hand will always remain handy in my future research. I would also like to acknowledge the guidance I received from Dr. Kavita Vemuri in the research collaborations we had during my time here.

I would also like to extend my sincere acknowledgement to IHub-Data for awarding me the prestigious one year research MS fellowship for pursuing a very crucial section of my overall research, without which my thesis would not have panned out the way I desired. The work that I carried out during the tenure of the fellowship plays a significant role in providing a viable solution to my thesis problem.

It is not just the supervisors and the mentors who enable you to fulfil your research ambitions, but also your peers who provide unconditional support and unwavering belief that ultimately push you to strive for the best. I want to thank my research collaborators C. Rajashekar Reddy and Ayush Kumar

Dwivedi for their help and support. I would also like to give a shout out to my friends Nayan, Ashutosh, Shubham, Gaurav and others. I would also like to thank my labmates Ayu, Spanddhana, Ruchi, Ishan and everyone else for creating a very lively and productive environment. I would like to thank the batch of MS students enrolled in 2019 for making my past few years memorable, in spite of the melancholia caused by the COVID-19 pandemic. I convey my best wishes to the new researchers who joined our research group for their future endeavors. Kindly accept my sincere apologies in case if I am missing out any names.

Lastly, all I have to say is that no amount of words of acknowledgement could completely suffice the unfathomable gratitude I have for my parents. They have been my pillars of support and encouragement. Everything that I am today has been possible because of their support, love, and blessings. Thanks for making my dreams come true, even if that means sacrificing yours. I know the road thus far has not been easy and I cannot say for sure that the future would not have any more challenges and struggles in store for us, but I can assure you that you have made your son capable enough to tackle whatever curve ball life can throw. I hope I made you proud and I continue to do so.

Abstract

Air quality monitoring and analysis of a given location can be challenging without being able to determine the factors influencing its behavior. Stationary air pollution monitoring systems mostly account for temporal variations in few locations due to the traditional choice of sparse deployment. In addition, the data from these monitoring stations may not be easily accessible. Internet of things (IoT) has allowed deployment of cost-efficient networks of smart devices with an ease in connectivity to the internet enabling easier air pollution monitoring.

This thesis mainly proposes two test cases that highlight prominent factors that deeply influence the pollution pattern in a location. In the first test case, a temporal variation analysis is performed on two networks of pollution monitoring systems to understand the effects of implementing a nationwide lockdown during the COVID-19 pandemic. The first network comprises of the costly and bulky air quality monitoring stations of Central Pollution Control Board (CPCB) sparsely deployed in Hyderabad city and the second network comprises of the smaller and low-cost IoT nodes densely deployed inside the IIIT-H campus. Differential analyses were conducted on the data to understand the effects of lockdown on Particulate Matter (PM) levels by factoring in the yearly and seasonal trends, followed by Welch's t-test to check whether the PM values have changed w.r.t. values in pre-lockdown period. Lastly, Pearson pair-wise correlation coefficient is estimated between PM and temperature values to show the effect of temperature changes on the PM values irrespective of lockdown. This test case established the effects of human-centric factors such as vehicular and industrial emissions, commercial activities and commotion on ambient PM variations as a sudden drop in the anthropogenic activities during the nationwide lockdown has caused a decline in PM levels.

In the second test case, a novel methodology to analyze a moving object database is proposed by performing a case study on mobile IoT data collecting Particulate Matter across a road stretch of India, and look into the neighboring spatial and anthropogenic factors such as human activities, settlement patterns and vegetation profile corresponding to each geo-location of the PM data. The thematic interactions of spatial and anthropogenic factors of each location with the corresponding ambient PM levels resulted in a factor-based data structure that highlights the PM distribution mapped to each factor. This case study not only enabled to showcase the influence of spatial and anthropogenic factors of a location in influencing its ambient PM levels, but also provided a use case for handling mobile object databases, a challenging issue prevalent amidst the GIS community. Both the test cases effectively demonstrated

the contribution of different surrounding factors in affecting the PM values, thus providing a gateway for accurately modeling the ambient air quality data based on key parameters.

Contents

List of Figures

List of Tables

List of Abbreviations

AQM	Air Quality Monitor
API	Application Program Interface
AI	Artificial Intelligence
TCP/IP	Transmission Control Protocol/Internet Protocol
BLE	Bluetooth low energy
CPCB	Central Pollution Control Board
COVID-19	Corona Virus Disease - 2019
CPS	Cyber Physical Systems
ICT	Information and Communications Technology
IIIT-H	International Institute of Information Technology - Hyderabad
IEEE	Institute of Electrical and Electronics Engineer
IIoT	Industrial Internet of Things
IoT	Internet of Things
NDVI	Normalized Difference Vegetation Index
LoRa	Long Range
LPWAN	Low-Power Wide Area Network
LTE	Long Term Evolution
ML	Machine Learning
DL	Deep Learning
M2M	Machine to Machine
NB-IoT	Narrowband IoT
PM	Particulate Matter
QOS	Quality of Service
RH	Relative Humidity
SSN	Static Sensor Networks
VoLTE	Voice over Long Term Evolution
WAN	Wide Area Network
WiFi	Wireless Fidelity
WSN	Wireless Sensor Network
WHO	World Health Organaisation

Chapter 1

Introduction

1.1 Motivation

Over the past few years, the ease in complex computation and high speed internet connectivity has resulted in the emergence of an innovative technology known as the Internet of Things (IoT) (**cite**). In a nutshell, IoT is the concept of connecting any thing to the Internet and to other connected devices. The IoT is a giant network of connected things and people – all of which collect and share data about the way they are used and about the environment around them. That includes an extraordinary number of objects of all shapes and sizes – from smart home systems, which monitor and/or control home attributes such as lighting, temperature, entertainment systems and appliances, to self-driving cars, whose complex sensors detect objects in their path, to wearable fitness devices that measure your heart rate and the number of steps you've taken that day, then use that information to suggest exercise plans tailored to you. There are even connected footballs that can track how far and fast they are thrown and record those statistics via an app for future training purposes [1].

Air pollution has been a cause of severe concern since ages due to its hazardous effects on human health and the environment. It has been a burning agenda amongst global climate change advocates. There are several air pollutants in the atmosphere, out of which Particulate Matter (PM) has been identified as the most crucial contributor to air pollution [2, 3]. Long term exposure to PM may increase the chances of severe respiratory and cardiovascular illness. Considering the diverse range of IoT applications, one it is extremely pertinent to incorporate novel IoT-based air pollution monitoring and analysis techniques. Numerous research have been conducted in the past to accurately monitor and evaluate the impact of air pollution using IoT-based solutions [4, 5, 6]. Most studies emphasize on temporal variation of pollutants to understand ambient air quality. As the temporal variation analysis is performed on regular day-to-day air pollutant concentration, these studies exclusively account for the influence of meteorological factors, but fail to account for any direct correlation of any anthropogenic or spatial factor over the pollutant levels. This thesis is designed to address and evaluate the influence of anthropogenic and spatial factors on pollutant (in this case, PM) concentration levels by cumulatively presenting two sets of experimental analysis carried out on low-cost IoT nodes. In the first experiment, the temporal variation of PM is studied to evaluate the effects of the nationwide COVID-19 lockdown that caused a sudden drop in human activities, thus determining a correlation between PM and anthropogenic activities. Whereas, in the second experiment, a spatial variation analysis is performed on the PM data collected over a diverse range of geo-locations across the country. Two networks of stationary IoT nodes were employed in the first experiment, whereas a mobile IoT node was used in the second experiment to collect the PM data. Both these experiments not only evaluated the influence of spatial and human-centric factors in surrounding PM variation, but also promoted the need for a hybrid IoT-based solution comprising of stationary and mobile IoT nodes for capturing optimum spatio-temporal resolution of PM variation.

1.2 Summary of Contributions

The main contributions of this thesis are split among two chapters:

- **Chapter 3**
  - In the study presented in this chapter, the focus is on quantifying the change in PM concentration in Hyderabad city due to the implementation of nationwide COVID-19 lockdown.
  - Two datasets have been used for this analysis. The first dataset (data for one and half years) is from CPCB stations in the city [7], while the second dataset (data for six months) is from the dense IoT network of PM monitors deployed in the educational campus of IIITH in Gachibowli region of Hyderabad, a bustling IT and financial area [6].
  - Differential analyses are done on the data to understand the effect of lockdown on PM values by factoring in the yearly and seasonal variations.
  - The Welch's t-test is carried out to test whether the PM values have changed with respect to values in pre-lockdown time period.
  - Pearson pair-wise correlation coefficient is estimated between PM and temperature values to show the effect of temperature changes on the PM values irrespective of lockdown.
- **Chapter 4**
  - This study proposes a novel methodology of analyzing mobile IoT data by performing spatial and anthropogenic factor-based thematic interactions with it to retrieve interesting patterns that account for the data variation.
  - Data is collected for a large range of geo-locations across the country that showcases a diverse spatial variation during a limited time period using a mobile IoT node. The data primarily comprises of PM concentration and vegetation cover corresponding to said geo-locations.

- Few road stretches within different routes were identified in the spatially varying PM profile where PM levels showed significantly higher values (spikes). The anthropogenic activities and land use features of these road stretches provided sufficient evidence to further investigate the influence of spatial and anthropogenic factors over ambient PM levels, thus raising a need for formulating a generic framework that could explain the spatial variation of the PM data collected over the course of the entire journey of the mobile IoT node.
- A factor-based data structure is proposed comprising the minimum, mean, maximum and standard deviation of PM values obtained by performing thematic interactions between different spatial and anthropogenic factors corresponding to each geo-location. This type of factor analysis is performed to understand the influence of spatial and anthropogenic factors over PM variation.

1.3 Organization of Thesis

The remainder of this thesis is organized as follows:

- *Chapter 2* provides an overview of IoT, types of IoT nodes, application of IoT in different areas and verticals and lastly the use of IoT in monitoring and analysis of air pollution data.
- *Chapter 3* gives an analysis of PM variation in Hyderabad city during COVID-19 lockdown.
- *Chapter 4* discusses a case study on PM variation across India using spatial factor-based analysis of data obtained from a mobile IoT node.
- *Chapter 5* serves as the conclusion of the thesis.

Chapter 2

IoT : A Brief Overview

This chapter gives an overview about an IoT node and its components. It is followed by a brief description of its two basic classifications: stationary and mobile IoT nodes. Subsequently, application and use-case of IoT nodes in air pollution monitoring are discussed. Lastly, the major challenges faced by traditional IoT-based air pollution monitoring are presented. This chapter provides only a brief introduction, and interested readers can read further details from various interesting books and articles on IoT like [8, 9, 10, 11, 12]

2.1 Introduction

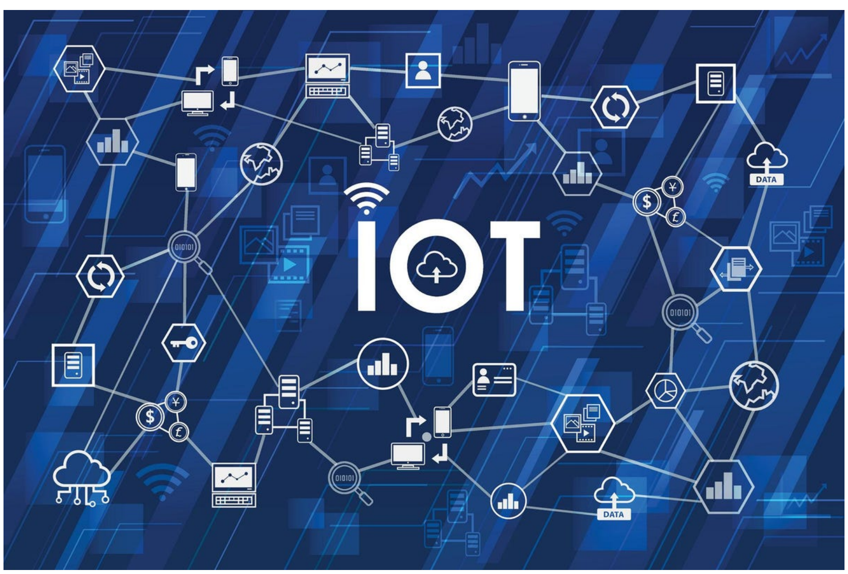


Figure 2.1: Internet of Things

IoT is about interconnecting embedded systems, bringing together evolving technologies such as sensors, actuators, high performance computing, wireless connectivity, and cloud storage/internet, as seen in Fig.2.1. These connected embedded systems are independent microcontroller-based computers that use sensors to collect data. These IoT systems (also known as IoT nodes) are networked together usually by a wireless protocol such as WiFi, Bluetooth, 802.11.4, or a custom communication system. The networking protocol is selected based on the distribution of nodes and the amount of data to be collected. This data is sent over the network to the main hub or computer. This main computer collects and analyzes the data, storing it in memory and even making system decisions based on the results of the analysis.

To understand what truly defines the expression **Internet of Things**, one needs to go back in time, in the year 1999, when the expression was first coined by Kevin Ashton, the British technology pioneer who co-founded the Auto-ID Center at the Massachusetts Institute of Technology, which created a global standard system for RFID and other sensors [13]. According to Ashton,

> "*The Internet of things is about empowering computers...so they can see, hear and smell the world for themselves*"

Few eminent agencies and organizations have given their own definitions of IoT:

- Gartner Research [14] defines it as *the network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment.*
- United Nations International Telecommunication Union [15] defines it as *a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.*

2.2 Major components

A typical IoT system can be divided majorly into five major components as shown in Fig. 2.2 and listed as thing/device, gateway, cloud, analytics and user interface [16].

2.2.1 Smart devices and sensors – Device connectivity

Devices and sensors are the components of the device connectivity layer. These smart sensors are continuously collecting data from the environment and transmit the information to the next layer. Latest techniques in the semiconductor technology is capable of producing micro smart sensors for various applications. Some of the common sensors are: temperature sensors and thermostats, pressure sensors, humidity / moisture sensors, light intensity detectors, proximity detection, RFID tags, etc. Most of the modern smart devices and sensors can be connected to low power wireless networks like Wi-Fi, ZigBee, Bluetooth, Z-wave, LoRAWAN etc. Each of these wireless technologies has its own pros and cons in terms of power, data transfer rate and overall efficiency. Developments in the low power, low cost wireless transmitting devices are promising in the area of IoT due to its long battery life and efficiency.

Figure 2.2: IoT Components

[16]

2.2.2 Gateway

IoT gateway manages the bidirectional data traffic between different networks and protocols. Another function of gateway is to translate different network protocols and make sure interoperability of the connected devices and sensors. Gateways can be configured to perform pre-processing of the collected data from thousands of sensors locally before transmitting it to the next stage. IoT gateway offers certain level of security for the network and transmitted data with higher order encryption techniques. It acts as a middle layer between devices and cloud to protect the system from malicious attacks and unauthorized access.

2.2.3 Cloud

IoT creates massive data from devices, applications and users which has to be managed in an efficient way. IoT cloud offers tools to collect, process, manage and store huge amount of data in real time. Industries and services can easily access these data remotely and make critical decisions when necessary. Basically, IoT cloud is a sophisticated high performance network of servers optimized to perform high speed data processing of billions of devices, traffic management and deliver accurate analytics. Distributed database management systems are one of the most important components of IoT cloud. Cloud system integrates billions of devices, sensors, gateways, protocols, data storage and provides predictive analytics. Companies use these analytics data for improvement of products and services, preventive measures for certain steps and build their new business model accurately.

2.2.4 Analytics

Analytics is the process of converting analog data from billions of smart devices and sensors into useful insights which can be interpreted and used for detailed analysis. Smart analytics solutions are inevitable for IoT system for management and improvement of the entire system. One of the major advantages of an efficient IoT system is real time smart analytics which helps engineers to find out irregularities in the collected data and act fast to prevent an undesired scenario. Service providers can prepare for further steps if the information is collected accurately at the right time. Big enterprises use the massive data collected from IoT devices and utilize the insights for their future business opportunities. Careful analysis will help organizations to predict trends in the market and plan ahead for a successful implementation. Information is very significant in any business model and predictive analysis ensures success in concerned area of business line.

2.2.5 User interface

User interfaces are the visible, tangible part of the IoT system which can be accessible by users. Designers will have to make sure a well designed user interface for minimum effort for users and encourage more interactions. Modern technology offers much interactive design to ease complex tasks into simple touch panels controls. Multicolor touch panels have replaced hard switches in our household appliances and the trend is increasing for almost every smart home devices. User interface design has higher significance in today's competitive market, it often determines the user whether to choose a particular device or appliance. Users will be interested to buy new devices or smart gadgets if it is very user friendly and compatible with common wireless standards.

2.3 IoT applications

There are several applications and use cases where IoT plays a defining role. Some of the applications and use cases of IoT in a real-world scenario are listed below:

- **Smart cities:** The issues related to smart cities include air pollution, distribution of water, traffic congestion, security [17, 18, 19]. Air pollution monitoring has been one of the major applications of IoT in the past few years. Smart and green buildings are an integral part of smart cities with building automation and smart HVAC control systems for energy-efficient activities. IoT networks help improve the quality of service (QoS) and improve the efficiency of the energy management system. It would also help to ease the control, and proportional power distribution to the population across the cities. In India, the smart city mission was launched in 2015, where hundred cities and towns are selected for improving the quality of life. The mission aims to drive economic growth and improve the quality of life of people by enabling local area development and harnessing technology, especially technology that leads to smart outcomes, and ensures that these cities are livable, inclusive, sustainable, and have thriving economies.

- **Connected Industry**: The IIoT, being the basis for the Industry 4.0 and Smart Factory, provides connectivity for smart factories, machines, industrial infrastructure, management systems, and more to streamline business operations, creating intelligent, self-optimizing industrial equipment and facilities [20, 21]. A number of issues related to the management of equipment and resources, the security and safety of people – can be addressed with innovative IoT solutions. The Industrial IoT represents a fast-growing field of application due to its exceptionally low power capabilities perfectly suited for the industrial automation world, enabling innovative services for improving efficiency, reliability, and availability of industrial processes and products.

- **Health care:** IoT devices are reshaping the way we receive healthcare [22, 23, 24], and are equipping not just normal people in monitoring and tracking their health in everyday life, but also companies-sponsors, doctors, and patients with innovative insights and analytics. The data obtained from the IoT devices help healthcare, pharma, and life science companies to make better decisions and gain a competitive advantage. The advancements in IoT and its continuous coordination with Pharma and the Healthcare Industries will launch the evolution of real-time monitoring and treatment of diseases. The pandemic of COVID-19 in recent times has accelerated the adoption of IoT in healthcare. In times of physical distancing, IoT has allowed remote diagnosis as well as treatment of patients. During the ongoing COVID-19 crisis, IoT has played a pivotal part in properly monitoring patients who are virus-infected through devices and intertwined networks.

- **Wearables:** Wearables such as smartwatches, smart bands, fitbits have started to dominate the market. These consist of sensors embedded inside them to sense different body parameters, such as heart rate, activity tracking. The data collected can be processed and used to calculate derived parameters like calories burnt, sleep tracking, heart rate monitoring, etc. [25, 26]. Smart wearables collect and analyze data, and in some scenarios, make a smart decision and provide a response to the user and are finding more and more applications in our daily life.

- **Autonomous vehicles:** The automobile industry has been there for a long time and has always been evolving ever since [27, 28, 29]. But, the major transformation is happening now from vehicles being driven by humans to being driven by themselves autonomously. Today's cars have already been extensively connected. The automobile industry is on the brink of a revolution to move to the self-driving automobile industry, and the driving force behind this is the fast-developing technology, the IoT. It will transform the automobile industry, and at the same time, the automobile industry will provide a big boost to it. The potential and the prospects of this technology are astonishing.

- **Smart grids:** The grid refers to the electric grid, a network of transmission lines, substations, transformers, and more that deliver electricity from the power plant to your home or business. The Smart grid represents an unprecedented opportunity to move the energy industry into a new era of reliability, availability, and efficiency that will contribute to our economic and environmental health. Smart grid strives to improve the energy consumption of buildings. IoT networks are used to improve the quality-of-service and increase the efficiency of the energy management system for millions of buildings connected over different cities. It also enables ease of control and proportional power distribution to the population across the cities [30].

2.4 IoT for Air Pollution Monitoring: Discussion and Challenges

2.4.1 Discussion

Air pollution is a global challenge for governments, regulators, city administrators and citizens. Many governments are investing multi-billion dollar sums in policies and solutions to improve air quality and they are empowering cities to tackle air pollution locally. In order to implement effective policies and interventions there is an increasing focus on understanding the levels and causes of air pollution. Today, air quality monitoring is performed by large, expensive scientific instruments permanently installed and professionally maintained, at a relatively small number of fixed locations. For example Hyderabad has around six monitoring stations [7]. This makes it difficult for citizens to understand the levels of pollution they experience in their daily lives, as the monitoring data is not available in real time and is very sparse. Advances in sensors, IoT platforms and mobile communications technologies have led to the emergence of smaller, portable, low cost, mobile-enabled sensors that can measure and report air quality in near real time [31, 32]. The IoT connected devices sense the environment several times a minute and typically deliver a one minute average value to a connected analytical solution, creating an opportunity for research bodies to offer air pollution monitoring and control services that deliver dynamic, local information to stakeholders. Big data capabilities, such as analytical and machine learning, can then be applied to this data and related data sets, such as weather and traffic, to understand the causes and fluctuations in air pollution [33].

2.4.2 Challenges

Although the emergence of IoT in air quality monitoring and analysis has revolutionized the environmental engineering space, but there are still a lot of challenges that are required to be addressed in order to better utilize IoT-based solutions for garnering deeper insights on PM variation in the ambient atmosphere. Most of the air pollution research primarily account for temporal variations of pollution data and attempt to estimate or forecast pollutant concentrations [34, 35], but there is still scarcity in studies that provide a holistic understanding of pollution behavior. This is because in order to come up with possible explanations for pollution behavior, it is of utmost importance to determine the factors that influence ambient pollutant concentration patterns. A thorough understanding of factors affecting pollution levels can be achieved by exclusively studying temporal and spatial variations of pollutant concentrations by employing stationary and mobile IoT deployment strategies respectively.

IoT deployment strategies need to evolve further to incorporate accountability of most, if not all, factors that have a direct or indirect influence over PM variation. Without having a deeper understanding of these factors and how important of a role they have in deciding the PM concentration levels of a location at any given time, it is difficult for any IoT-based air pollution monitoring solution to build a smart environment. Most of the recent studies focus on implementing ML/DL algorithms in historical PM data of a given location and estimating future PM values [36, 37, 38, 39]. Although these studies account for meteorological factors like temperature, humidity, wind speed, wind direction, etc. of the location, they fail to account for spatial factors like vegetation profile and anthropogenic factors such as traffic patterns, industrial/commercial activities, settlement patterns, etc. In order to have a holistic spatio-temporal understanding of PM behavior, it is pertinent to evaluate the contribution of as many factors affecting PM as possible. Most recent research such as [40] have started working on building models that account for spatio-temporal variations throughout an entire city. Although [40] has been successful in implementing novel ConvLSTM model for interpolating and predicting air pollution data across a city by keeping into consideration few spatial and temporal features of the pollution data such as traffic volume, average driving speed and meteorological data, it fails to consider the influence of many other factors. On top of that, considering only city-wide data for proper understanding of PM behavior would not be diverse enough to highlight conclusive results.

Considering the importance of the different set of capabilities that stationary and mobile IoT nodes have to offer, it makes sense to incorporate both while formulating a solution for efficient air pollution monitoring and analysis. Now, in order to implement a hybrid model of air pollution monitoring comprising of stationary and mobile IoT nodes, it becomes pertinent to perform separate case studies on stationary and mobile IoT nodes which could show how both of these nodes have different challenges and know-hows of data collection and unique methods of data analysis. These exclusive methods of analysis used by stationary and mobile IoT nodes respectively provide two different vantage points to understand PM behavior in depth.

2.5 IoT Node Types

IoT nodes can be classified into the following two major types based on their deployment strategies:

2.5.1 Stationary

Stationary IoT node is a type of sensor node which is static in nature, i.e. the sensor node is placed in a specific location to sense data of that particular location. These kind of nodes are positioned in a location for a long period of time in order to collect data and monitor the data behavior of that particular location. After collecting data from location using a stationary IoT node for a significant time period, temporal variation analysis is usually performed to understand the effects of seasonality on the collected data. Therefore, stationary IoT nodes are the go-to option for temporal behavior understanding of data collected from a given location.One such example of stationary IoT node deployment could be the IoT network of PM monitors deployed in IIITH [6]. In [6], a dense deployment of IoT enabled low-cost sensor nodes is done. For this, total nine low-cost IoT nodes monitoring PM are deployed in a small educational campus in Hyderabad city. A web-based dashboard website is developed to easily monitor the real-time PM values. Another example of stationary IoT deployment, having relatively sparse deployment than [6] would be the six pollution monitoring stations of the CPCB, Hyderabad [7].

2.5.2 Mobile

Mobile IoT node is a type of sensor node which, as the name suggests, is mobile in nature, i.e. the sensor node is in motion due to its positioning on a mobile object (most commonly on terrestrial vehicles or aerial drones). They move from one location to another to sense data of different locations. These kind of nodes usually travel across different locations in order to collect different data values of different locations in a short period of time. Depending on the speed of the mobile object carrying the IoT node, mobile IoT nodes could cover a lot of locations within a short span of time, thus allowing to understand spatial variability of data over a diverse range of locations. Mapping of the mobile IoT data is the most popular and preliminary mode of visualization technique adopted during analysis. One example of mobile IoT deployment would be [41], where a black carbon (BC) measurement campaign was conducted along two fixed routes in Antwerp, Belgium using a bicycle equipped with portable BC monitor. Another good example of mobile IoT deployment would be the roadside case study to understand the spatial variability of air quality in Sydney, Australia [42]. The measurement campaign was conducted by mounting the PM monitoring device on top of a pram and collecting data over the chosen area of Randwick in Sydney, because it was also the subject area for an agent-based traffic model.

Chapter 3

IoT Network-Based Analysis of Variations in Particulate Matter due to COVID-19 Lockdown

This chapter discusses the analysis of PM variation across Hyderabad city during the COVID-19 lockdown period. An introduction to the study, followed by a brief overview on the IoT measurement networks used for collecting the dataset is presented, along with the analysis tools and techniques applied on the datasets, and lastly the analysis results are presented.

3.1 Introduction

A recent study has shown a correlation between PM (2.5 and 10) and mortality exposure due to the COVID-19 virus [43]. Localised and frequent monitoring of PM and other air pollutants is required for focused health advisories and intervention.

Internet of things (IoT) is the most preferred choice for air pollution monitoring due to its ability to sense and connect with the ambient surroundings and ease of interaction with users and other systems by employing an array of smart devices [33]. For example, Central Pollution Control Board (CPCB) has deployed six monitoring stations in Hyderabad measuring different pollution parameters linked to the Indian air quality index (AQI) including PM2.5, and PM10 [7]. These values are accessible to the general public through their website or downloaded by an APP through API. Although CPCB monitors are highly reliable, they are incredibly costly, and only a few can be deployed. For example, a large metropolitan city like Hyderabad spread across 650 km^2 has only six monitoring stations. This sparse deployment leads to the unavailability of pollution data at places of personal interest to the general public, such as residential areas, offices, and schools. This issue has paved the way for the low-cost but dense deployment of IoT networks for air pollution monitoring by researchers and institutions to understand local pollution. In our study a dense network of eight nodes in a small area of 66 acres was deployed in IIITH for monitoring PM2.5 and PM10 [6] as a pilot study. This study focuses on the monitoring of PM2.5 and PM10 parameters.

Significant contributors to the PM are vehicles, residential and industrial fuel burning, and dust [44, 2, 45]. After the outbreak of the corona pandemic in February 2020, governments worldwide have put several restrictions on human activities in public places as a measure to reduce the spread of the virus. In some countries, complete lockdown had been announced. The Indian Government announced a complete nationwide lockdown on March 24, 2020, which lasted till May 3, 2020. Due to the lockdown, the traffic, industrial, and other outdoor activities have dropped to a bare minimum. This has resulted in a significant drop in air pollution in several regions, including New Delhi, India [46], USA [47], and Europe [48]. The focus of this paper is on quantifying the effect of COVID-19 on the change of PM values in the southern Indian city of Hyderabad.

The specific contributions of this chapter are

- Two datasets have been used for this analysis. The first dataset (data for one and half years) is from CPCB stations in the city [7], while the second dataset (data for six months) is from the dense IoT network of PM monitors deployed in the educational campus of IIITH in Gachibowli region of Hyderabad, a bustling IT and financial area [6].
- Differential analyses are done on the data to understand the effect of lockdown on PM values by factoring in the yearly and seasonal variations.
- The Welch's t-test is carried out to test whether the PM values have changed with respect to values in pre-lockdown time period.
- Pearson pair-wise correlation coefficient is estimated between PM and temperature values to show the effect of temperature changes on the PM values irrespective of lockdown.

Unlike the studies in [47, 46, 48], which only rely on the data publicly available for few sparsely placed nodes in a city, the contribution in this study is the data from a dense deployment of low-cost sensors. Also, the seasonal variation and temperature effect have not been considered in the studies while calculating the impact of COVID-19 lockdown on the PM levels.

3.2 IoT Network Measurements

In this section, the measurement network for CPCB is explained first, followed by the details on the measurement network for the IIIT-H network.

3.2.1 CPCB nodes

Fig.3.1 shows the six pollution monitoring stations deployed in the Hyderabad city by CPCB [7][49]. The node in Zoo Park was not functional for most of the measurement period, and hence, it has not been considered for analysis. Each of these stations uses PM sensors with a resolution of 0.5 $\mu g\,m^{-3}$, a precision of ±2 $\mu g\,m^{-3}$ (1-hour average), and accuracy of ±1% [50]. The CPCB website provide hourly

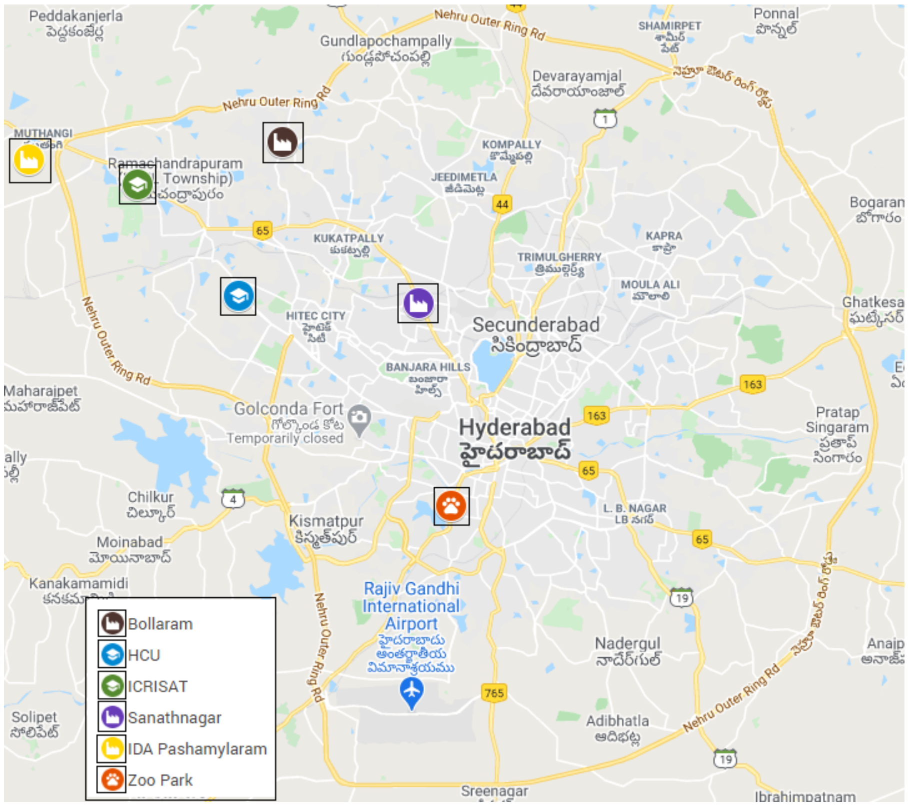


Figure 3.1: Locations of CPCB pollution monitoring stations in Hyderabad.

averaged values for PM2.5 and PM10, both in $\mu g\,m^{-3}$, in the *csv* format. For CPCB nodes, the data is collected from January 1, 2019, to June 30, 2020.

3.2.2 IIIT-H nodes

Fig. 3.2 shows the deployment of PM monitoring nodes developed and deployed in IIIT-H [6][49]. The primary reason for selecting IIIT-H campus for deploying these nodes and performing this study was a certain spatial and anthropogenic advantage that the campus provided. The campus has high vehicular traffic density on a regular basis outside, while it also possessed green patches of land and a dense population of students inhabiting inside, thus helping in highlighting the spatial and anthropogenic causes that might influence ambient pollution levels. The objective of dense deployment (eight nodes in 66 acres) was to measure PM value with high spatio-temporal resolution. Out of these eight nodes, Node8 was not functional for most of the measurement period, and hence it has not been considered for analysis. Each node has a Nova PM SDS011 sensor, with a resolution of 0.3 $\mu g\,m^{-3}$ and a relative error of max. ±15% (±10 $\mu g\,m^{-3}$)[51]. Each node is connected to the internet through a WiFi connection to either IIIT-H routers or a 4G based WiFi router, as mentioned in [6]. The IIIT-H nodes' data have been collected using the REST API from ThingSpeak server[52], where the nodes dump the sensor data. Each node provides PM2.5 in $\mu g\,m^{-3}$, PM10 in $\mu g\,m^{-3}$, temperature in °C and relative humidity

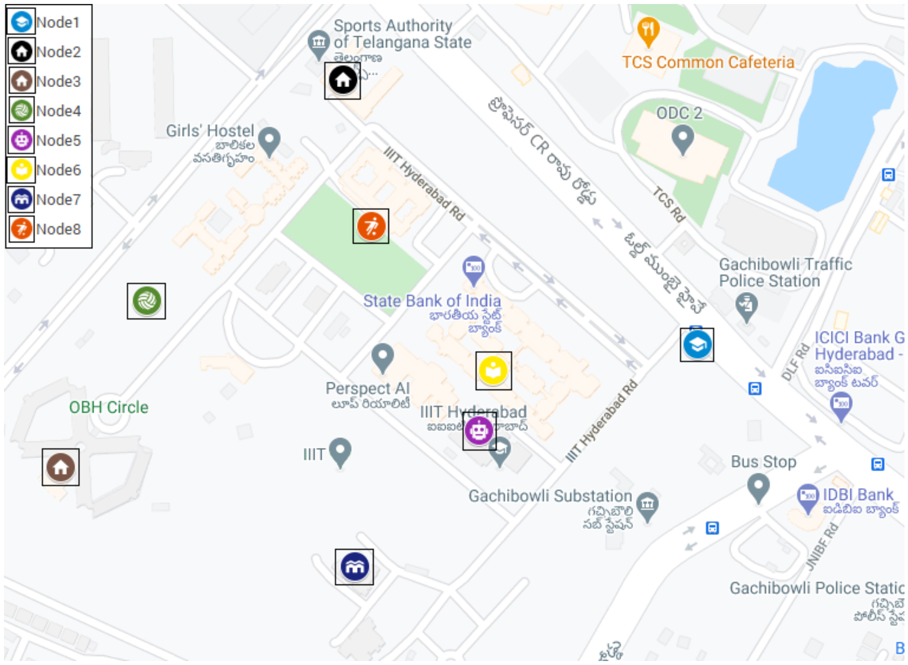


Figure 3.2: Deployment of PM monitoring IoT network in the IIIT-H campus.

every 15 seconds with an additional network lag. For IIIT-H nodes, the data collection is ongoing since October 26, 2019, but the data considered for analysis in this paper is from January to June 2020.

3.3 Data Processing Techniques

3.3.1 Data pre-processing

The raw data collected from the CPCB and IIIT-H IoT networks must be pre-processed before any analysis can be done. In this paper, we have considered two methods. The first one involves removing null data points, while the second one consists of the removal of any possible outliers. This is necessary to avoid any unnecessary deviations caused by extreme values. For removing outliers, Z-score [53] is used.

3.3.2 Analysis techniques

In this section, four analysis techniques considered in this paper are briefly presented: averaging of data, evaluating seasonal and yearly variations, and significance test (Welch's t-test).

3.3.2.1 Averaging of data

Generally, a moving average is used for analyzing data. Averaging helps in smoothing out short-term variations and reveal the long term trends or patterns. In this paper, hourly averaged data is used for finding statistical quantities such as mean, variance, and correlation between different quantities. Weekly average plots have been used for easy visualization, while monthly average values have been used for the change analysis.

3.3.2.2 Change analysis

For performing change analysis, February 2020 and April 2020 have been considered to represent the pre-lockdown (normal) and total lockdown periods. The months of March and May in 2020 were not considered for the same because these were transitional months, i.e., lockdown happened during the last week of March 2020, and the Government began relaxing the total lockdown rules during May 2020. This made February and April the ideal months to highlight the contrast between pre-lockdown and complete lockdown and establish a comparative understanding.

To understand the seasonal variations, the change for the monthly average PM values in April w.r.t. February for the years 2019 and 2020 respectively are considered. For the yearly trend variations, the relative change for the monthly average PM values in 2020 w.r.t. 2019 for February and April, respectively, are considered.

In this paper, the definition of change Δ and relative change R between monthly average PM values of month $m1$ of year $y1$ w.r.t. the reference month $m0$ of the year $y0$ is given by

$$\Delta = M_{m1,y1} - M_{m0,y0} \tag{3.1}$$

and

$$R(\%) = \frac{\Delta}{M_{m0,y0}} \times 100, \tag{3.2}$$

where $M_{m,y}$ is the monthly average PM value for the month m in the year y.

3.3.2.3 t-test

The t test is used to decide if there is a significant difference between the means of two groups. Let us denote the first group or dataset $\mathcal{A}$ as the PM values change for April 2019 w.r.t. February 2019. This corresponds to seasonal change during normal times (without lockdown). Similarly, let us denote the dataset $\mathcal{B}$ containing the PM values change for April 2020 w.r.t. February 2020. This dataset refers to the seasonal change during lockdown (abnormal). Using these two data sets, the two hypotheses corresponding to the t-test are

$$\begin{aligned} H_n &: \mathcal{A} == \mathcal{B} \\ H_a &: \mathcal{A} \,!= \mathcal{B} \end{aligned} \tag{3.3}$$

where H_n is called the null hypothesis (signifies no effect of lockdown on the PM levels) while H_a is called the alternative hypothesis (signifies the effect of lockdown on the PM levels).

Given that the two sets might have different means and variances, we employ Welch's t-test, where the test statistic is given by [54]

$$t = \frac{\bar{\mathcal{B}} - \bar{\mathcal{A}}}{s}. \tag{3.4}$$

Here $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ are means of the sets $\mathcal{A}$ and $\mathcal{B}$, respectively and s is the scaling parameter given by

$$s = \sqrt{\frac{s_{\mathcal{A}}^2}{N_{\mathcal{A}}} + \frac{s_{\mathcal{B}}^2}{N_{\mathcal{B}}}} \tag{3.5}$$

with $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$ are the number of samples and $s_{\mathcal{A}}$ and $s_{\mathcal{B}}$ are standard deviations for the sets $\mathcal{A}$ and $\mathcal{B}$, respectively.

The performance parameter in t-tests is p-value, which is the probability of correctly deciding the null hypothesis. A very small p-value means that such an extreme observed outcome would be improbable under the null hypothesis, implying the null hypothesis can be rejected.

3.4 Analysis and Results

In this section, results are presented in two parts. Data from CPCB nodes is analyzed in the first part, followed by the analysis of data from the IIIT-H IoT network.

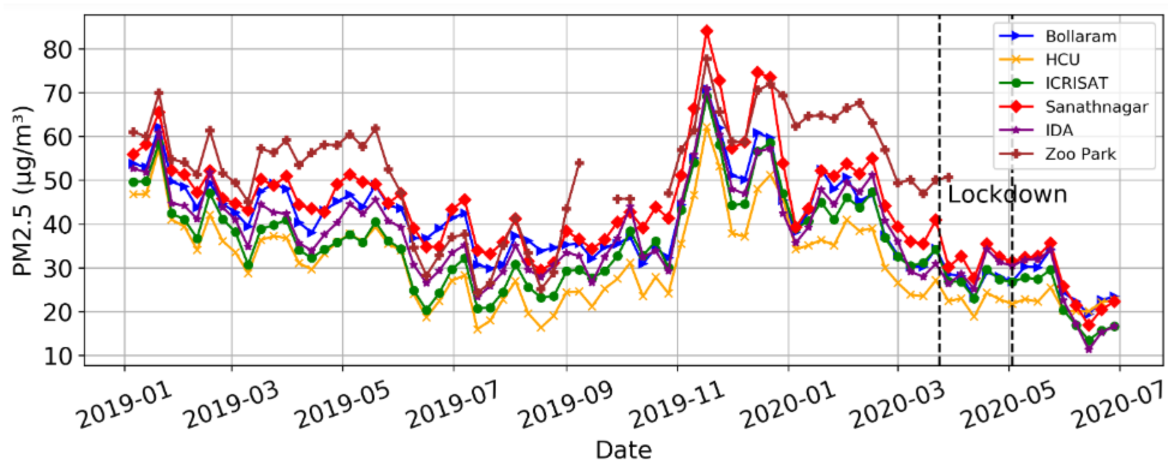
3.4.1 CPCB nodes

The CPCB data is analyzed by first plotting the central moving averages of PM values. Next, the change analysis is done taking seasonal and yearly variations into account, followed by the t-test.

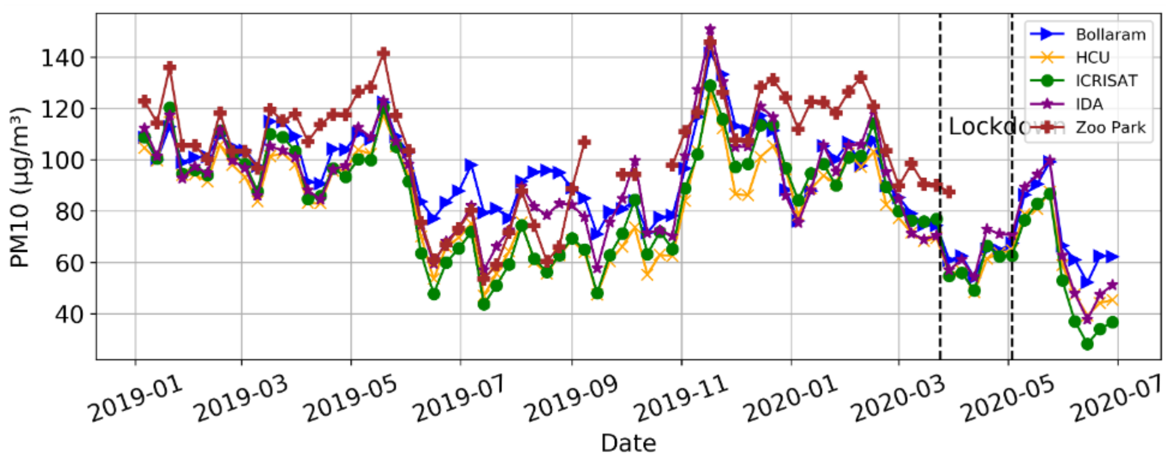
3.4.1.1 Averaging of data

Fig. 3.3 presents the central moving averages over the window of three weeks for PM2.5 and PM10 values across the CPCB stations through January 1, 2019, to June 30, 2020. It can be observed that the PM values are at a peak during winters (November-January). The PM values start decreasing with an increase in temperatures in February and keep dropping through summer before hitting the lowest values in Monsoon (June-September) because of rains. The values start increasing with the onset of winter in October and again hit a peak in November. It can be also observed from Fig. 3.3 that both the PM values are the lowest for the nodes at the HCU and ICRISAT. This is expected as HCU and ICRISAT are green institute campuses while other nodes are in industrial areas. Note that the graph of weekly average PM10 values for Sanathnagar station is missing in Fig. 3.3b due to data non-availability. It can be also observed that the data at the Zoo Park node is missing for February 2020 and hence it has been omitted in further analysis.

Table 3.1 shows the monthly average values of PM2.5 and PM10 for February 2019, April 2019, February 2020, and April 2020. Similar to Fig. 3.3, it can be observed that the PM values in general decrease going from February to April irrespective of the lockdown, though the difference is significantly



(a) PM2.5 ($\mu g\, m^{-3}$)



(b) PM10 ($\mu g\, m^{-3}$)

Figure 3.3: Central moving averages of PM values for the CPCB stations in Hyderabad with window length of 3 weeks.

high for the months in the year 2020. Therefore, to understand the effect of lockdown, we need to take into account yearly and seasonal variations.

3.4.1.2 Yearly variations

Table 3.2 shows the yearly variation in the PM values regarding change and relative change. It can be seen that the PM2.5 values have slightly reduced in February, going from 2019 to 2020 in Bollaram and HCU while they have increased marginally for other stations. Similarly, PM10 values in February have decreased somewhat for Bollaram, HCU, and ICRISAT while showing a slight increase for other stations. For example, the relative change in PM2.5 varies from -7.07% to 5.11% , while for PM10, the relative change varies from -6.88% to 3.28%. However, the PM values have only decreased and that too significantly in April across the two years for all stations. For example, the relative change in PM2.5

Table 3.1: Monthly average PM values for CPCB stations

(a) PM2.5 ($\mu g\,m^{-3}$)

Station	**Feb2019**	**Apr2019**	**Feb2020**	**Apr2020**
Bollaram	46.15	40.12	42.88	27.75
HCU	37.48	31.23	35.13	23.06
ICRISAT	39.94	33.69	41.98	28.52
Sanathnagar	49.18	43.61	49.28	33.64
IDA	44.76	34.63	46.34	32.01

(b) PM10 ($\mu g\,m^{-3}$)

Station	**Feb2019**	**Apr2019**	**Feb2020**	**Apr2020**
Bollaram	105.74	94.65	98.46	61.94
HCU	98.79	86.67	93.33	59.31
ICRISAT	100.08	89.20	99.35	62.91
IDA	102.85	87.34	106.23	69.23

varies from -30.83% to -7.58% while for PM10, the relative change varies from -20.73% to -34.55%. Thus, the effect of lockdown on reduction of pollution is visible in the yearly trends.

3.4.1.3 Seasonal variations

Table 3.3 shows the seasonal variations, i.e., change in the values of PM going from February to April in the same year. Here, the year 2019 serves as a reference year for normal times. It can be seen that the PM values decrease from February to April for both the years. This is as expected with the increase in temperatures and can also be seen from Fig.3.3. The relative decrease in the seasonal variation in 2019 ranges from 11.32% to 22.61% for PM2.5 and from 10.48% to 15.07% for PM10. However, the relative changes in the seasonal variation in 2020 are consistently show the relative decrease upwards of 30%. This shows the effect of lockdown on PM values, that is a decrease.

3.4.1.4 t-test

All the results presented till now are based on observing the differences in the monthly averages. To be sure that there is indeed change in the PM values, we consider t-test, which takes into account the standard deviations of the two data sets. The data for set $\mathcal{A}$ belongs to the change in the hourly averaged PM values for April 2019 w.r.t. February 2019. The data for set $\mathcal{B}$ belongs to the change in the hourly averaged PM values for April 2020 w.r.t. February 2020. Table 3.4 presents the t-test results for the same. It can be seen that the null hypothesis is rejected with more than 99.99% confidence for all the stations for both PM2.5 and PM10 values validating the decrease in the PM values observed in the Tables 3.2 and 3.3 caused because of the lockdown. Only exception is IDA which is also rejecting the null hypothesis with 95% confidence. The CPCB stations at the industrial locations of Bollaram and Sanathnagar show the biggest variation followed by the greener institute HCU and ICRISAT. IDA, which is an industrial area at the outskirt of the city, is showing the least variation.

Table 3.2: Yearly variation (2020 w.r.t. 2019)

(a) PM2.5

Station	**Change (Feb)** ($\mu g\,m^{-3}$)	**Relative Change (Feb) (%)**	**Change (Apr)** ($\mu g\,m^{-3}$)	**Relative Change (Apr) (%)**
Bollaram	-3.263	-7.07	-12.371	-30.83
HCU	-2.347	-6.26	-8.168	-26.14
ICRISAT	2.042	5.11	-5.168	-15.33
Sanathnagar	0.099	0.20	-9.972	-22.86
IDA	1.584	3.53	-2.626	-7.58

(b) PM10

Station	**Change (Feb)** ($\mu g\,m^{-3}$)	**Relative Change (Feb) (%)**	**Change (Apr)** ($\mu g\,m^{-3}$)	**Relative Change (Apr) (%)**
Bollaram	-7.279	-6.88	-32.705	-34.55
HCU	-5.458	-5.52	-27.358	-31.56
ICRISAT	-0.730	-0.72	-26.296	-29.47
IDA	3.382	3.28	-18.108	-20.73

3.4.2 IIIT-H IoT network

The IIIT-H IoT network data is analyzed in three parts: averaging of data, Pearson r correlation, and seasonal variation analysis. As the historical data of 2019 is not available, analyses corresponding to yearly variations and t-test have not been carried out. Pearson's r correlation analysis has been performed to establish the correlation between temperature variation and PM values change.

3.4.2.1 Averaging of data

Fig. 3.4 presents the weekly average variation in PM2.5 and PM10 from January to July 2020. The weekly average PM values are least for Node3 (located at the most interior part of the campus) and highest for Node1 (situated in the campus's main entrance gate). Of the seven nodes, Node4 shows the steepest descent, especially after the lockdown happened, which implies that the most significant decrease was recorded in Node4. The slope for the graphs of Node3 and Node7 remained almost constant during the lockdown period, which denotes that these two nodes did not record any significant change during the lockdown. It can be seen that data from Node8 is missing for the entire February and has not been considered for further analysis.

Table 3.3: Seasonal variation (April w.r.t. February)

(a) PM2.5

Station	**Change (2019)** ($\mu g\, m^{-3}$)	**Relative Change (2019) (%)**	**Change (2020)** ($\mu g\, m^{-3}$)	**Relative Change (2020) (%)**
Bollaram	-6.028	-13.06	-15.137	-35.29
HCU	-6.247	-16.66	-12.068	-34.34
ICRISAT	-6.248	-15.64	-13.459	-32.05
Sanathnagar	-5.570	-11.32	-15.641	-31.73
IDA	-10.123	-22.61	-14.334	-30.92

(b) PM10

Station	**Change (2019)** ($\mu g\, m^{-3}$)	**Relative Change (2019) (%)**	**Change (2020)** ($\mu g\, m^{-3}$)	**Relative Change (2020) (%)**
Bollaram	-11.090	-10.48	-36.517	-37.08
HCU	-12.121	-12.26	-34.022	-36.45
ICRISAT	-10.871	-10.86	-36.437	-36.67
IDA	-15.507	-15.07	-36.998	-34.82

Table 3.4: t-test analysis

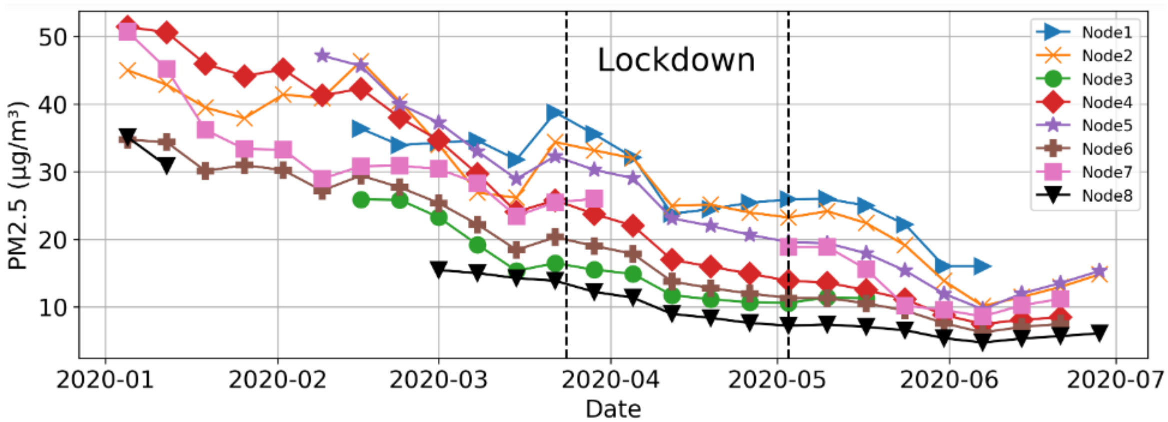
Station	**t-value (PM2.5)**	**p-value (PM2.5)**	**t-value (PM10)**	**p-value (PM10)**
Bollaram	-8.043	2.38e-15	-9.597	5.60e-21
HCU	-3.862	0.00011	-7.331	4.88e-13
ICRISAT	-5.455	6.26e-08	-7.050	3.62e-12
Sanathnagar	-7.663	4.81e-14	-	-
IDA	-2.039	0.0417	-5.817	8.60e-09

3.4.2.2 Pearson's r correlation analysis

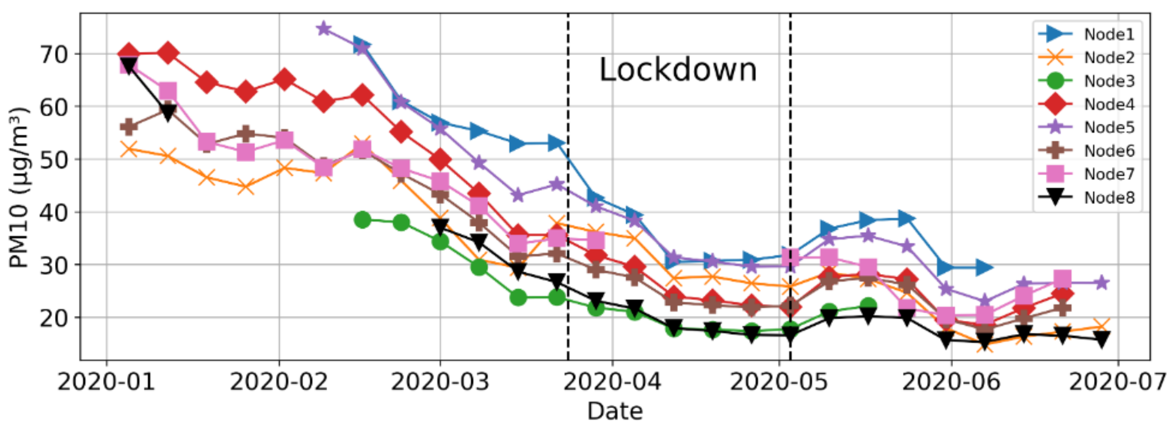
From Table 3.5, it can be seen that there is a strong negative correlation between temperature and PM values across all the nodes. It demonstrates that the PM values, even in normal times, decrease with increase in temperature values going from winter to summer. This shows that although there is a significant decrease in the PM values during the lockdown, not all PM values reduction can be attributed to the lockdown. Therefore the seasonal and yearly variations along with t-test analysis are essential.

3.4.2.3 Seasonal Variation

Table 3.6 shows the comparison of the decrease of PM values in April 2020 with respect to February 2020 in terms of concentration Change ($\mu g\, m^{-3}$) and Relative Change (%). It has been observed that the decrease is noticeably different for both PM10 and PM2.5 for the seven node locations. The relative



(a) PM2.5 ($\mu g\, m^{-3}$)



(b) PM10 ($\mu g\, m^{-3}$)

Figure 3.4: Central moving average plot with window length of 3 weeks for IIIT-H IoT nodes.

change for 7 Nodes in the IIIT-H campus ranged from -31.3% to -55.05% in case of PM2.5 and -39.26% to -56.94% for PM10. The difference in the values for the seven nodes is significant and considering that all nodes are in a small area of 66 acres shows the variation in the level of human activities in a small campus and the need for a dense deployment of PM monitoring nodes for localizing the events of cause for the PM values.

Table 3.6b show that almost every node deployed in the campus of IIIT-H show a consistent relative change (%) of at least 39% to maximum of 56% which is a variation of 16%. But in Table 3.6a it can be observed that the decrease in PM2.5 values show a similar trend as the CPCB stations deployed across the city, with few nodes showing a comparatively slighter decrease in terms of change ($\mu g\, m^{-3}$) even though the relative change (%) is higher.The relative change is from a minimum of around 31% to a maximum of 55% which amounts to a difference in around 24% of relative change in the campus. Consideration of the sources of PM2.5 in the surroundings of the respective nodes can give a possible explanation of the low or high drop of the PM2.5 values measured by the individual nodes in these locations.

Table 3.5: Pearson's r correlation coefficient analysis for Jan 2020 to Mar 2020 variation (Temperature vs PM)

Pearson's r	**Node1**	**Node2**	**Node3**	**Node4**	**Node5**	**Node6**	**Node7**
PM2.5	-0.67	-0.58	-0.80	-0.81	-0.37	-0.85	-0.74
PM10	-0.86	-0.64	-0.87	-0.80	-0.42	-0.83	-0.74

Table 3.6: Nodewise monthly average, change and relative change values for PM in IIIT nodes.

(a) PM2.5

Node ID	**Feb 2020** ($\mu g\,m^{-3}$)	**Apr 2020** ($\mu g\,m^{-3}$)	**Change** ($\mu g\,m^{-3}$)	**Relative Change (%)**
Node1	40.28	27.21	-13.07	-32.44
Node2	42.58	27.38	-15.20	-35.71
Node3	25.882	12.75	-13.12	-50.72
Node4	41.78	18.78	**-23.00**	**-55.05**
Node5	36.74	25.24	-11.50	-31.3
Node6	31.79	15.19	-16.60	-52.22
Node7	28.25	17.73	-10.51	-37.22

(b) PM10

Node ID	**Feb 2020** ($\mu g\,m^{-3}$)	**Apr 2020** ($\mu g\,m^{-3}$)	**Change** ($\mu g\,m^{-3}$)	**Relative Change (%)**
Node1	73.132	33.843	**-39.289**	-53.72
Node2	49.578	30.112	-19.466	-39.26
Node3	37.838	18.901	-18.937	-50.05
Node4	60.772	26.171	-34.601	**-56.94**
Node5	58.371	33.803	-24.568	-42.09
Node6	54.989	24.365	-30.624	-55.69
Node7	47.001	25.134	-21.867	-46.52

- Node1 and Node2, even though being the closest to the vehicular pollution from the six-lane road in front of the campus, do not show a substantial decrease in terms of change ($\mu g\,m^{-3}$) and relative change (%) because of the construction and repair of the same road taken up by the Greater Hyderabad Municipal Corporation (GHMC, the civic body that oversees the city of Hyderabad) during lockdown [55], which provided uninterrupted access to the repair and construction of the roads and flyovers. These construction hampered the expected decrease of the PM2.5 values even without the significant source of pollutants, vehicles being absent during the lockdown.

- Node4 shows the maximum decrease in both changes of -23.0 $\mu g\,m^{-3}$ and relative change of -55.05 % during the lockdown. The area surrounding Node4 is an online food and other e-commerce order delivery point and is busy round the clock with delivery vehicles before the lockdown and has a football ground nearby, typically used for sports activities before the lockdown. During the lockdown period, the area around Node4 is completely idle with no above mentioned human activities.

- Node6 shows the second-highest relative change of -52.22 % and change of -16.60 $\mu g\, m^{-3}$ during the lockdown. The node is located where the sources contributing to the PM2.5 values such as massive scale and round the clock fuel burning for cooking in the canteen just beside and dispersion of settled dust by people moving around during the rush hours of classes throughout the day have been entirely devoid due to the suspension of classes, and a closure of the canteen during the lockdown period, contributed to a higher decrease in PM2.5 values.

- Node3, Node5, and Node7 show the least amount of change around 10 $\mu g\, m^{-3}$. Node3 offers a higher relative change value of -50.72% even with a change of just -13.12 $\mu g\, m^{-3}$ as the initial value 25.882 $\mu g\, m^{-3}$ of this node is lowest the other nodes in February 2019 and are in locations where the human activities are much less even before the lockdown than the other nodes. Node3 and Node7 are in residential areas for students and faculty away from main activities, and Node5 is placed behind the research block, with low vehicle movement or construction activity,away from the significant PM values sources like vehicular pollution or large scale fuel burning for cooking. As the PM2.5 sources in these locations are already low even before the lockdown, the complete shutdown during the lockdown did not show substantial improvement in PM2.5 values.

The above points explain how PM2.5 decreased due to lockdown in only the human activities' locations (Node4 and Node6). Even with the primary source of PM2.5 values, i.e., vehicular pollution was nearly nil in Node1 and Node2, other human activities like construction of the road contributed to the PM2.5 values. Nodes with already lower values of PM2.5 even before the lockdown - Node3, Node5, and Node7 - did not show substantial change due to lockdown. Note that the variation across the different nodes for PM10 is not high when compared to PM2.5 which is a similar pattern seen in CPCB data. Hence, any explanation for PM10 has not been provided.

3.5 Conclusion

In this study, by factoring in the yearly and seasonal trend analysis and applying t-test on the CPCB data, it has been demonstrated that there is a consistent decrease in the PM values across all the nodes because of the COVID-19 lockdown. A similar trend is observed for the data obtained for a smaller area of IIIT-H campus using the IoT sensor network deployed. However, the correlation analysis has shown a strong negative correlation between the temperature and the PM values demonstrating that not all the decrease in the PM values is because of lockdown. Moreover, the considerable variation in the effect of lockdown on the reduction in PM values in a small IIIT-H campus shows the importance of dense deployment for PM monitoring, identification of localised sources of pollution and the contribution of each source to the values.

Chapter 4

Spatial Factor Analysis of Mobile IoT Data : A Case Study on Particulate Matter across India

This chapter provides the motivation for monitoring air pollution and using IoT as an enabler for it, followed by global initiatives around the world for tackling air pollution, conventional monitoring sensor networks, low-cost sensors for air pollution monitoring, and a thorough survey of various existing IoT air pollution monitoring networks around the world.

4.1 Introduction

Stationary IoT deployment have been traditionally used for air pollution monitoring and analysis. Although stationary IoT nodes enable in better understanding the ambient pollution levels during continuous monitoring and account for temporal variations of pollution, they lack in understanding cross-sectional changes. Therefore, in order to consider the effects of spatial variations on ambient pollution levels, it is essential to incorporate an effective mobile IoT deployment strategy. Recent studies have shown several challenges in mobile IoT data handling, storage and analysis for air pollution data due to the involvement of several variables and factors that play a vital role in ambient pollution levels. Studies like [42, 41] have proposed innovative methodologies to monitor and analyze mobile PM data to account for spatial variability. The role of vegetation cover in PM variation has also been cited in many research studies over the years [56, 57].

For this study, a mobile IoT node was deployed that collected PM data over a diverse range of geo-locations across India, thus creating a moving object database. This moving object database is then subjected to thematic interactions with spatial and anthropogenic factors to retrieve interesting patterns in the data. By doing so, a novel methodology of data structuring and analysis is conceptualized that accounts for the spatial variations in the PM data. The specific contributions of this paper are three-fold. Firstly, it's the PM data collection for a large range of geo-locations across the country that showcases a diverse spatial variation during a limited time period using a novel self-made IoT node. Secondly, the analysis of the spatially varying data to identify PM value peaks and their corresponding

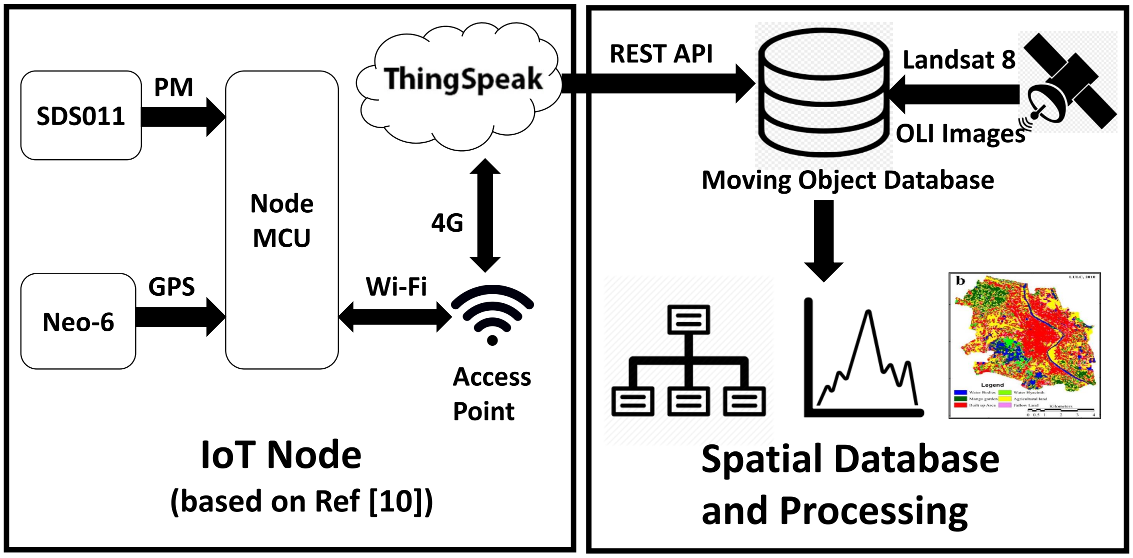


Figure 4.1: Architecture Diagram

activities happening near the location. And lastly, the proposed factor-based data structure containing the minimum, mean, maximum and standard deviation of PM values obtained as a result from interactions between different vegetation and anthropogenic factors corresponding to each geo-location.

The rest of the paper is organized as follows. Section II describes the hardware specifications of the mobile IoT node and the adopted data processing techniques. Section III discusses the analysis and results obtained, while Section IV concludes the paper.

4.2 Hardware Specifications and Data Processing

4.2.1 Hardware Specifications

The IoT node has primarily four integral components: an ESP8266 micro-controller unit (NodeMCU)[58], with an inbuilt Wi-Fi module, an SDS011 Nova PM sensor[59] to measure PM2.5 and PM10, a Neo-6 GPS sensor module[60] used to fetch geo-location values, and a portable broadband router to provide internet connectivity for offloading the data to the cloud server[52]. Although the sampling period for the node was 30 seconds, there were missing data points during collection due to internet connectivity issues and the GPS sensor unable to capture Lat-Long values for certain locations. Fig.4.2 shows the mobile IoT node used in this study[6].

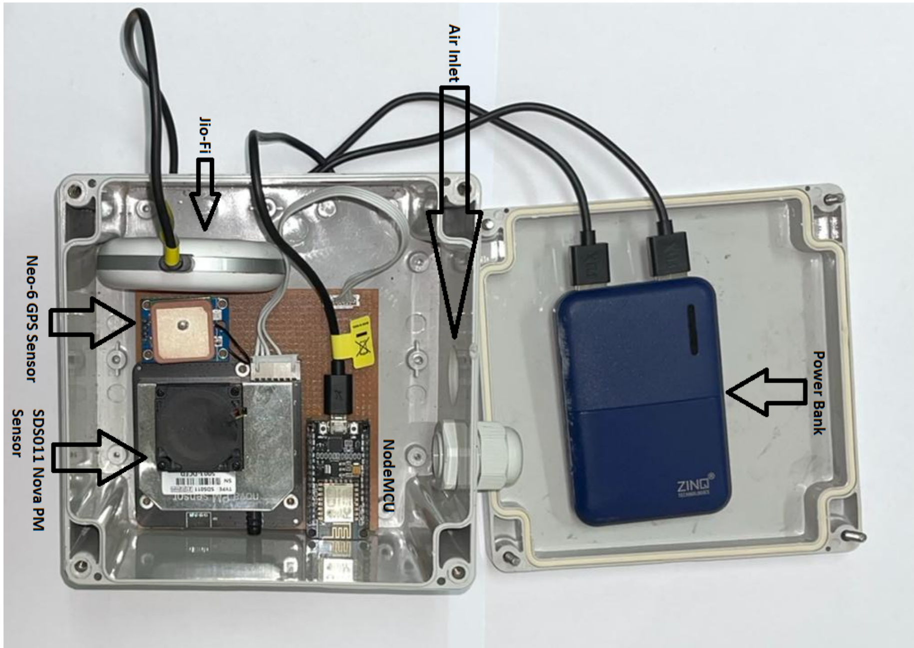


Figure 4.2: IoT Node (Hardware)

4.2.2 Data Collection

4.2.2.1 Data Collection Campaign

The IoT node was carried inside a car with open windows and travelled along the route (as seen in Fig.4.3) in order to collect mobile data. The PM data and the corresponding Latitude-Longitude (Lat-Long) values of a particular geo-location were collected using the REST API from ThingSpeak server[52]. The first phase of the mobile node journey took place between 24th October, 2020 to 25th October, 2020. The journey began from Hyderabad, Telangana and ended in Agra, Uttar Pradesh, covering 1021 km (approx.). Along the journey, the mobile node traversed through the states Telangana, Maharashtra, Madhya Pradesh, Rajasthan and Uttar Pradesh. The second phase was longer than the first and took place between 6th December, 2020 to 10th December, 2020, from Agra, Uttar Pradesh and it ended in Bengaluru, Karnataka, covering 1655 km (approx.). During the journey, the mobile node traversed through the states Uttar Pradesh, Rajasthan, Gujarat, Maharashtra and Karnataka.

4.2.2.2 Data Preparation

A total of 8482 PM data points were collected, out of which only 4653 PM data points were further processed based on completeness of the observations of the IoT node (collected every 30 seconds) in two stages of the journey. Out of the 4653 points, 4274 points were in National Highways (NH), which is what we considered for our study. Out of these 4274 points, 1205 points were from the first phase of the journey, whereas 3069 points were from the second phase of the journey. Subsequently, the data

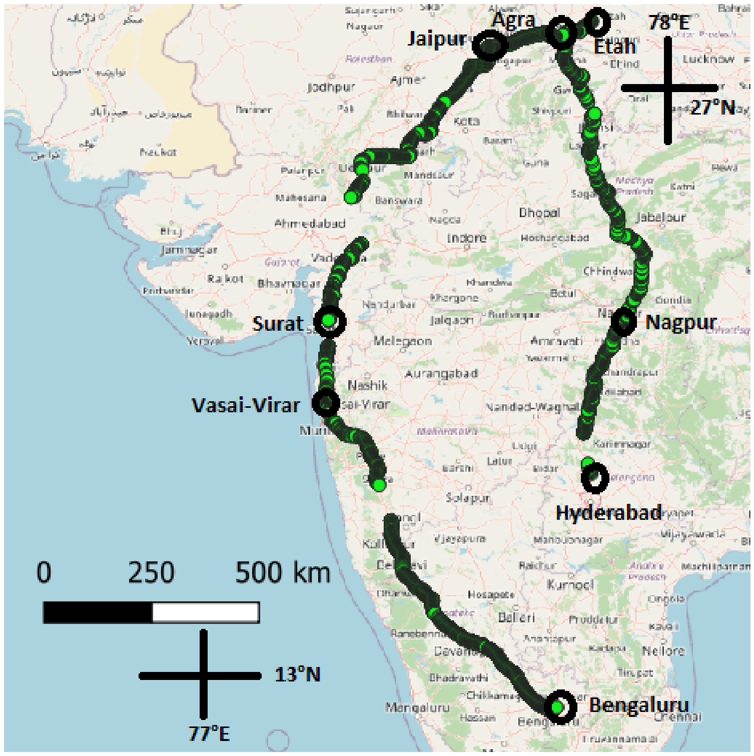


Figure 4.3: Mobile IoT Node Data

is subjected to preliminary data cleaning process. The data cleaning process involves removal of null data points and outliers. In order to perform outlier removal, a standard Z-score based outlier removal technique is employed [53].

4.2.3 Data Processing Techniques

4.2.3.1 Factor-based data structure

The PM10 data is classified into two ranges, 0<PM10<100 and PM10>100, as 100 $\mu\mathrm{g\,m^{-3}}$ is the satisfactory level threshold [7]. Furthermore, the data points are categorized corresponding to their possibly associated anthropogenic **factors** in their vicinity, which are labelled on the basis of: (a) Act or Non-Act: The data points are checked if it is located near any activity (Act) zone or not. An activity zone is considered to be any place which has significant industrial activities (industry zones), commercial activities (market/commercial areas) or traffic activities (toll plazas/road junctions) (b) Settlement patterns (Rural/Semi-Urban/Urban): After Act/NonAct classification, it is further checked if the data point is in a rural (ActR/NonActR), semi-urban (ActS/NonActS) or urban (ActU/NonActU) area.

After classifying the PM10 ranges into respective categories such as NH, Act/NonAct, ActR/NonActR, ActS/NonActS and ActU/NonActU, the PM10 ranges are represented in the form of number of data points (Count), minimum (Min), mean (Mean) and maximum value (Max) along with its standard deviation (STD) each of the respective PM10 range.

4.2.3.2 Normalized Difference Vegetation Index (NDVI)

This is the most commonly used vegetation index for observing greenery[61]. In general, Healthy vegetation is good absorber of electromagnetic spectrum in visible reason. Chlorophyll contains in a greeneries highly absorbs Blue (0.4 - 0.5 μm) and Red (0.6 - 0.7 μm) spectrum and reflects Green (0.5 – 0.6 μm) spectrum. Therefore, our eye perceives healthy vegetation as green. Healthy plants have high reflectance in Near Infrared (NIR) between 0.7 to 1.3 μm. This is primarily due to internal structure of plant leaves. Due to high reflectance in NIR and high absorption in Red spectrum, these two bands are used to calculate NDVI.

$$NDVI = \frac{NIR – Red}{NIR + Red}, \tag{4.1}$$

where **NIR** is the reflectance in the Near Infrared band and **Red** is the reflectance in the Red band. In this study, Landsat 8 OLI images [62] were used for NDVI calculation, hence NIR is Band 5 and Red is Band 4. The NDVI value varies from -1 to 1. Higher the value of NDVI, higher or denser the greenery. In this study, NDVI values corresponding to each data point of the mobile node have been calculated for buffer regions of 500 m and 1 km respectively. The NDVI values are accounted for in settlement pattern analysis in the Factor-based data structure in order to associate the role of vegetation, if any, in the PM10 level variations of the specific settlement region. In this study, the NDVI values have been categorized into two ranges: 0-0.2 (Low vegetation), 0.2-0.4 (Moderate to high vegetation). Fig.4.2 represents the architecture diagram of the study comprising of two blocks (IoT Node and Spatial Database and Processing) that interact with each other, thus providing a generic framework to collect and analyze mobile data.

4.3 Analysis and Results

In this section, results are presented in two parts. In the first part, spatial variation of PM concentration is discussed to highlight regions with peak PM values, followed by factor-based data structuring of the entire mobile IoT data.

4.3.1 Peak PM10 values throughout the mobile node journey

Few road stretches within different routes were identified in the overall mobile node journey in Fig.4.3 that showed significantly higher PM10 levels (PM10>100) during the journey as seen from Fig.4.4:

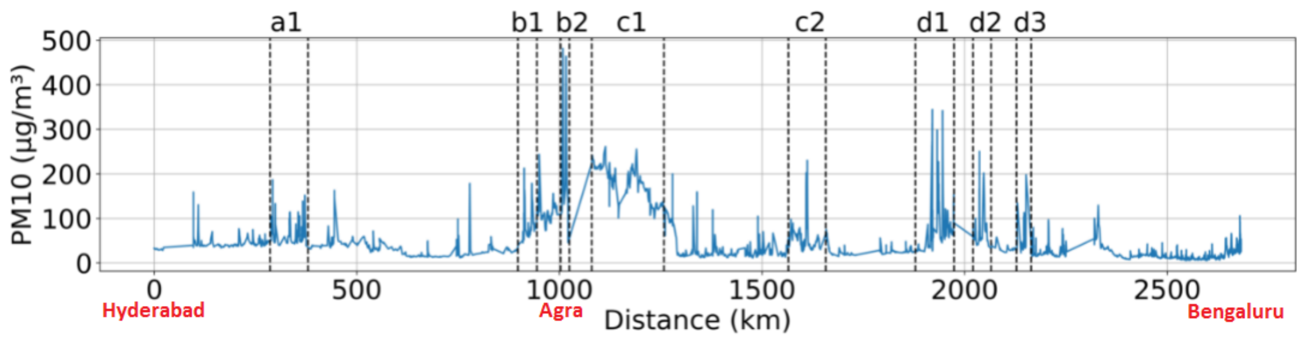


Figure 4.4: PM10 Plot of Mobile IoT Node Journey

- Hinganghat to Nagpur : This route had a stretch from 286-380 kms as seen in section **a1** in Fig.4.4 that witnessed PM10>100. Possible reason could be the on-going construction of highway as part of the Bharat Mala project[63].
- Gwalior to Agra : The stretches from 898-945 kms. and 1003-1025 kms. in this route as seen in sections **b1** and **b2** respectively in Fig.4.4 witnessed PM10>100, which could be attributed to high traffic and a lot of under-construction roads in the stretch.
- Etah to Jaipur : This route had two stretches from 1080-1258 kms and 1565-1657 kms. as seen in sections **c1** and **c2** respectively in Fig.4.4 having PM10>100. Both these stretches had very high traffic, accompanied by residual particle emissions from the nearby marble industries and factories.
- Bharuch to Surat and Valsad to Vapi: This route also had two stretches from 1879-1974 kms and 2021-2065 kms. as seen in sections **d1** and **d2** respectively in Fig.4.4 having PM10>100, which could be attributed to the moderate to high traffic experienced in this stretch and also the fact that this route falls under the Gujarat Industrial Development Corporation (GIDC) corridor, thus experiencing industrial aerosol emission on a regular basis.
- Vasai-Virar to Thane : This route had a stretch from 2128-2164 kms as seen in section **d3** in Fig.4.4 that witnessed PM10>100 points. Such PM behavior could be due to the extremely heavy traffic on account of this stretch being part of a highly urbanized route.

4.3.2 Peak PM2.5 values throughout the mobile node journey

4.3.3 Factor Analysis

The factor-based data structure presented in this paper can be used to represent PM values for the following ranges:

- PM10: PM10>100 $\mu g\, m^{-3}$ and 0<PM10<100 $\mu g\, m^{-3}$.

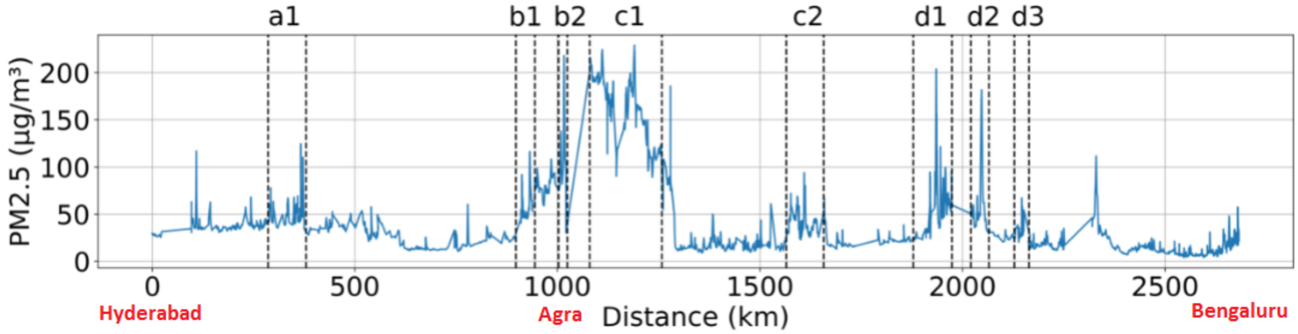


Figure 4.5: PM2.5 Plot of Mobile IoT Node Journey

- PM2.5: PM2.5>100 $\mu g\,m^{-3}$, 60<PM2.5<100 $\mu g\,m^{-3}$ and 0<PM2.5<60 $\mu g\,m^{-3}$.

4.3.3.1 PM10>100 $\mu g\,m^{-3}$

Following deductions can be made from Fig.4.6:

- Act vs NonAct: The PM10 levels of points that fall under Act are a little higher than Non-Act. This goes to show that activities around points in this PM10 range have limited influence PM10 levels and are largely local.

- ActR vs ActS vs ActU: Further looking at Act points, we can see that the PM10 levels in the urban region are the highest, followed by the semi-urban region. The PM10 levels in the rural region are significantly lesser than urban and semi-urban regions. Discussing about the vegetation profile of all the three settlement regions, the number of low vegetation points is significantly larger than moderate to high vegetation points. The gap between number of low and mod-high vegetation points doesn't change much as buffer size increases. Thus, vegetation of the three regions don't have any direct influence over the PM10 levels.

- NonActR vs NonActS vs NonActU: Checking into the NonAct points, we can see that the PM10 levels in the semi-urban region are the highest, followed by the urban region. In this scenario as well, the PM10 levels in the rural region are significantly less than semi-urban and urban regions. Here, the gap between rural and semi-urban/urban is even larger than Act points. Observing the vegetation profile, we find that the PM10 levels for low vegetation region is greater than that of mod-high vegetation region within the 500 m and 1 km buffer zone for urban points. For rural region, the low vegetation points show greater PM10 levels than mod-high vegetation points within 500 m buffer only. Therefore, it can be deduced that in NonActR and NonActU regions, vegetation plays a key role in influencing the PM10 levels (cannot say the same for NonActS as it has no mod-high vegetation points).

Table 4.1: PM10 distribution corresponding to Vegetation Profile for PM10>100 $\mu g\,m^{-3}$

(a) Non-ActR (500 m buffer)

NDVI Range	**Count**	**Min** ($\mu g\,m^{-3}$)	**Mean** ($\mu g\,m^{-3}$)	**Max** ($\mu g\,m^{-3}$)	**STD** ($\mu g\,m^{-3}$)
0-0.2	42	101	126.42	250.3	33.15
0.2-0.4	25	106.1	120.37	155.7	12.24

(b) Non-ActU (500 m buffer)

NDVI Range	**Count**	**Min** ($\mu g\,m^{-3}$)	**Mean** ($\mu g\,m^{-3}$)	**Max** ($\mu g\,m^{-3}$)	**STD** ($\mu g\,m^{-3}$)
0-0.2	44	101.5	175.73	344.1	76.45
0.2-0.4	9	100.1	165.64	481.4	121.92

(c) Non-ActU (1 km buffer)

NDVI Range	**Count**	**Min** ($\mu g\,m^{-3}$)	**Mean** ($\mu g\,m^{-3}$)	**Max** ($\mu g\,m^{-3}$)	**STD** ($\mu g\,m^{-3}$)
0-0.2	42	106	178.95	344.1	76.77
0.2-0.4	11	100.1	155.18	481.4	111.54

4.3.3.2 0<PM10<100 $\mu g\,m^{-3}$

Fig.4.6 provides some insightful points as stated below:

- Act vs NonAct: The mean PM10 levels of Act is almost twice as much as that of the NonAct. This goes to show that industrial and other economic activities around points in this PM10 range influence PM10 levels significantly.

- ActR vs ActS vs ActU: The settlement patterns of the Act points show that the PM10 levels in the semi-urban areas are followed closely by urban and rural areas with almost similar PM10 levels. This could be due to Industrial Zones (IDZ) outside the primarily residential rural and urban settlements. The vegetation profile within the 1 km buffer for the rural and semi-urban settlement region showed that the PM10 levels of low vegetation region were greater than that of the mod-high vegetation region. This goes to highlight the influence of vegetation in influencing the PM10 at low levels in ActR and ActS points.

- NonActR vs NonActS vs NonActU: The settlement pattern shows that 1790 out of 2577 NonAct points have rural settlement. That implies the PM10 level comparison amongst different settlement patterns here would yield inconclusive results. Also, the vegetation profile around these settlement regions have predominantly low vegetation upto 1 km buffer. Therefore, it is not possible to determine the influence of vegetation on low to moderate PM10 levels for these settlement patterns.

Some key takeaways from the above factor analysis of PM10 values can be noted as below:

Table 4.2: PM10 distribution corresponding to Vegetation Profile for 0<PM10<100 (1 km buffer)

(a) ActR

NDVI Range	**Count**	**Min** ($\mu g\, m^{-3}$)	**Mean** ($\mu g\, m^{-3}$)	**Max** ($\mu g\, m^{-3}$)	**STD** ($\mu g\, m^{-3}$)
0-0.2	262	14.4	53.22	100	18.38
0.2-0.4	92	21.8	45.97	89.8	15.08

(b) ActS

NDVI Range	**Count**	**Min** ($\mu g\, m^{-3}$)	**Mean** ($\mu g\, m^{-3}$)	**Max** ($\mu g\, m^{-3}$)	**STD** ($\mu g\, m^{-3}$)
0-0.2	167	14.6	58.12	99.7	23.64
0.2-0.4	67	23.4	45.96	99.6	18.24

- The influence of **Activity** around NH roadways on PM levels can be seen for both the ranges of PM10 values. It can be noted that the gap between PM10 values of Act and Non-Act regions is almost double in case of 0<PM10<100, whereas for PM10>100, the gap is not that significant.
- The PM10 levels were least for rural areas (irrespective of any activity around), whereas the PM10 levels were the highest for semi-urban areas having no activity around for both the ranges of PM10 values as they tend to lie on transport corridors.
- The influence of vegetation within 1km buffer over the PM levels in the range 0<PM10<100 can be noticed for Activity regions having rural and semi-urban settlements. The PM10 levels in the range PM10>100 are influenced by vegetation within 500 m buffer in the Non-Activity regions having rural and urban settlements, and the vegetation within 1 km buffer influences the PM10 levels in Non-Activity regions having urban settlements.

4.3.3.3 PM2.5>100 $\mu g\, m^{-3}$

Following points can be noted from Fig.4.7:

- Act vs NonAct: 287 out of 310 NH points fall under Activity region. Therefore, the PM2.5 level comparison between Activity and Non-activity region points would yield inconclusive results.
- ActR vs ActS vs ActU: Viewing into the settlement pattern of the Activity points around NH, it can be seen that the urban region has highest PM2.5 levels, followed closely by semi-urban region points, which exceeds the PM2.5 levels of rural regions quite significantly. Looking into the vegetation profile of the three regions, it can be seen that for rural and semi-urban regions, almost all points have Low vegetation. For urban region points within 1km buffer, the PM2.5 levels for Low vegetation points are greater than that of Moderate vegetation points (Almost all Low vegetation points within 500m buffer for urban region). Therefore, it can be said that vegetation influences a little in the PM2.5 levels of urban areas having activity around NH.

- NonActR vs NonActS vs NonActU: 19 out of 23 Non-activity region points of NH fall under semi-urban regions. Therefore, comparing the PM2.5 levels in the different settlement regions wouldn't yield any conclusive results. It can also be noted that the all these semi-urban points have low vegetation profile, thus inhibiting from making any deductions regarding influence of vegetation on the PM2.5 levels.

Table 4.3: PM2.5 distribution corresponding to Vegetation Profile for PM2.5>100 (ActU 1km buffer)

NDVI Range	**Count**	**Min** ($\mu g\, m^{-3}$)	**Mean** ($\mu g\, m^{-3}$)	**Max** ($\mu g\, m^{-3}$)	**STD** ($\mu g\, m^{-3}$)
0-0.2	69	101	166.82	228.8	38.17
0.2-0.4	19	100.1	155.8	190.9	27.81

4.3.3.4 60<PM2.5<100 $\mu g\, m^{-3}$

From Fig.4.8, the following conclusions can be drawn:

- Act vs NonAct: The PM2.5 levels of the Activity region around NH are greater than that of the Non-activity region. Therefore, it goes to show that Activity greatly influences the PM2.5 levels around NH.

- ActR vs ActS vs ActU: Unlike the Red zone, the yellow zone showed an opposite result in PM2.5 behavior across the different settlement regions. Here, it can be observed that the PM2.5 levels in the Activity regions of rural areas are the highest, closely followed by semi-urban points, whereas the urban points showcase the least PM2.5 levels. Viewing the vegetation profiles across the three settlement regions, all the three regions have mostly Low vegetation within 500m and 1km buffer. Although within 1km buffer for rural and urban points, the no. of Low and Mod vegetation points are comparable. But the PM2.5 levels in those Mod vegetation points are greater than that of the Low vegetation points. Therefore, it can be said that vegetation has no effect on PM2.5 levels in the Activity regions around NH.

- NonActR vs NonActS vs NonActU: The settlement pattern around the Non-activity regions of the NH points show predominantly rural and urban areas. It can be seen that the PM2.5 levels in the rural areas is greater than the PM2.5 levels of urban areas. The vegetation profiles of both these settlement regions show that the PM2.5 levels for Low vegetation points is greater than that of the Moderate vegetation points in urban areas, whereas in rural areas, the PM2.5 levels of Moderate vegetation points exceed the PM2.5 levels of Low vegetation points. Therefore, it can be conclusively said that vegetation influences the PM2.5 levels of urban NH points having no activity.

Table 4.4: PM2.5 distribution corresponding to Vegetation Profile for 60<PM2.5<100 (NonActU)

(a) 500m buffer

NDVI Range	**Count**	**Min** ($\mu g\, m^{-3}$)	**Mean** ($\mu g\, m^{-3}$)	**Max** ($\mu g\, m^{-3}$)	**STD** ($\mu g\, m^{-3}$)
0-0.2	81	60.3	71.19	91.8	9.54
0.2-0.4	14	62.8	70.05	77.9	5.26

(b) 1km buffer

NDVI Range	**Count**	**Min** ($\mu g\, m^{-3}$)	**Mean** ($\mu g\, m^{-3}$)	**Max** ($\mu g\, m^{-3}$)	**STD** ($\mu g\, m^{-3}$)
0-0.2	77	60.3	71.17	91.8	9.73
0.2-0.4	18	62.8	70.39	77.9	5.15

4.3.3.5 0<PM2.5<60 $\mu g\, m^{-3}$

Fig.4.8 lays out the following deductions:

- Act vs NonAct: The Activity region around NH showed PM2.5 levels more than twice the PM2.5 levels of Non-activity region points around NH. This highlights that the PM2.5 levels around NH are highly influenced by activities around the NH.
- ActR vs ActS vs ActU: The settlement patterns near the Activity region of NH showed similar levels of PM2.5 in semi-urban and rural areas, closely followed by the PM2.5 levels of urban areas. The vegetation profile around rural and semi-urban settlements within 1km buffer showed the PM2.5 levels in Low vegetation regions to be greater than that of Mod vegetation regions.
- NonActR vs NonActS vs NonActU: The settlement patterns near the non-activity regions of NH points were mostly rural, followed by urban settlements. Both these settlement regions showed similar PM2.5 levels. It can also be noted that the settlement regions were predominantly having Low vegetation within 1km buffer, thus not helping to understand any possible effect of vegetation on PM2.5 levels.

Some key takeaways from the above factor analysis of PM2.5 values can be noted as below:

- The influence of ”Activity” around NH roadways on PM2.5 levels can be seen for the range 0<PM2.5<100.
- The PM2.5 levels were highest for rural areas around NH roadways (irrespective of any activity around NH) for 60<PM2.5<100, whereas he PM2.5 levels were the least for urban areas around NH roadways (irrespective of any activity around NH) for 60<PM2.5<100.
- The influence of vegetation within 1km buffer over the PM2.5 range 0<PM2.5<60 can be noticed for Activity regions having rural and semi-urban settlements. For PM2.5>100, vegetation within 1km buffer influences the PM2.5 levels in Activity regions having urban settlements.

Table 4.5: PM2.5 distribution corresponding to Vegetation Profile for 0<PM2.5<60 (1km buffer)

(a) ActR

NDVI Range	Count	Min ($\mu g\, m^{-3}$)	Mean ($\mu g\, m^{-3}$)	Max ($\mu g\, m^{-3}$)	STD ($\mu g\, m^{-3}$)
0-0.2	240	13	39.93	59.9	11.75
0.2-0.4	92	19.5	37.15	59.8	10.45

(b) ActS

NDVI Range	Count	Min ($\mu g\, m^{-3}$)	Mean ($\mu g\, m^{-3}$)	Max ($\mu g\, m^{-3}$)	STD ($\mu g\, m^{-3}$)
0-0.2	139	11.4	40.67	59.9	14.37
0.2-0.4	65	20.4	36.59	59.6	10.41

4.4 Conclusion and Future Scope

In this study, by performing factor-based hierarchical data structuring, accompanied by section-wise PM10 profiling on the overall data collected using the mobile IoT node, helped assess impact of activities like commercial and industrial on the ambient PM10 levels of a location. It has also been observed that rural settlement regions experience lesser PM10 levels in general than semi-urban and urban regions. The study also highlighted the limited influence of vegetation cover around a location over its PM10 levels, indicating a more local effect. It has also been evidently observed in this study how economic and environmental factors are important in understanding moderate to high PM levels across a region. The framework proposed in this study for analyzing mobile IoT data can be extended for the analysis of moving object databases. The insights garnered from this case study accounted for the spatial variability of PM, which goes on to promote similar mobile IoT deployment strategies in areas that exclusively have stationary IoT deployment for PM monitoring.

0<PM10<100

Count	Min	Mean	Max	STD
3721	5.4	34.96	100	22.11

Act

Count	Min	Mean	Max	STD
1144	14.4	52.12	100	19.5

Act (Rural)

Count	Min	Mean	Max	STD
354	14.4	51.33	100	17.85

Act (Semi-Urban)

Count	Min	Mean	Max	STD
234	14.6	54.64	99.7	22.86

Act (Urban)

Count	Min	Mean	Max	STD
556	20.3	51.57	99.6	18.91

Non-Act

Count	Min	Mean	Max	STD
2577	5.4	27.34	99.8	18.66

Non-Act (Rural)

Count	Min	Mean	Max	STD
1790	5.4	26.47	99.8	18.16

Non-Act (Semi-Urban)

Count	Min	Mean	Max	STD
210	6.7	30.36	97	19.93

Non-Act (Urban)

Count	Min	Mean	Max	STD
577	5.9	28.93	99.4	19.51

PM10>100

Count	Min	Mean	Max	STD
552	100.1	160.56	481.4	53.6

Act

Count	Min	Mean	Max	STD
408	100.1	163.18	464	49.47

Act (Rural)

Count	Min	Mean	Max	STD
130	100.3	147.43	243.4	31.73

Act (Semi-Urban)

Count	Min	Mean	Max	STD
120	100.1	167.9	261.2	42.74

Act (Urban)

Count	Min	Mean	Max	STD
158	100.3	172.56	464	61.81

Non-Act

Count	Min	Mean	Max	STD
144	100.1	153.15	481.4	63.5

Non-Act (Rural)

Count	Min	Mean	Max	STD
67	101	124.16	250.3	27.31

Non-Act (Semi-Urban)

Count	Min	Mean	Max	STD
24	107.9	188.01	243.9	45.6

Non-Act (Urban)

Count	Min	Mean	Max	STD
53	100.1	174.02	481.4	84.47

Figure 4.6: Factor-based Hierarchical Data Structures (PM10)

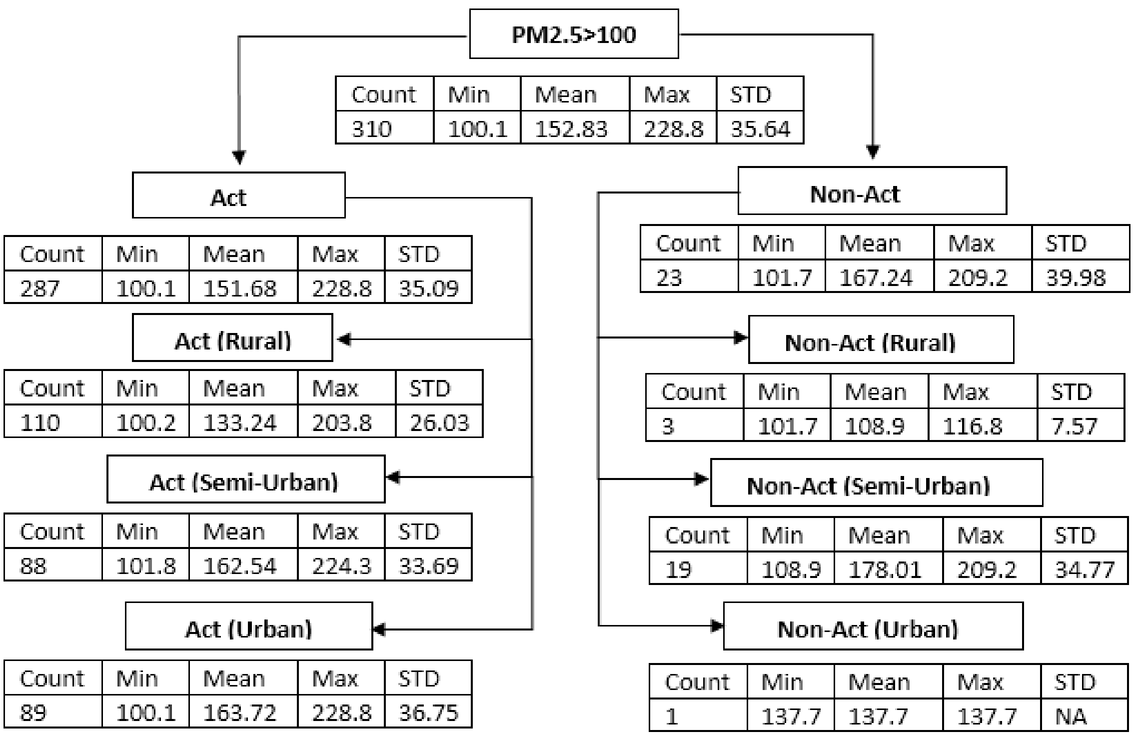


Figure 4.7: Factor-based Hierarchical Data Structure (PM2.5>100)

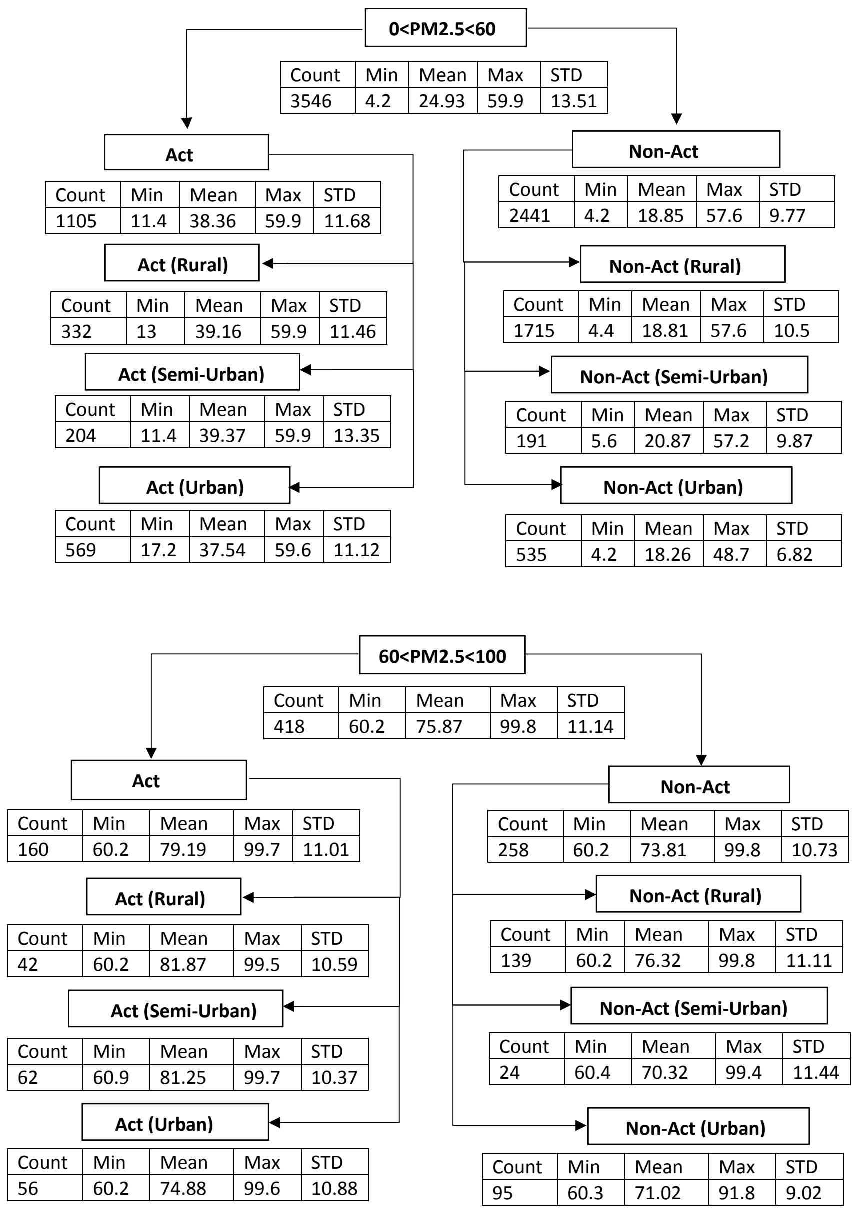


Figure 4.8: Factor-based Hierarchical Data Structures (0<PM2.5<60 and 60<PM2.5<100)

Chapter 5

Concluding Remarks

5.1 Conclusions

The research carried out in this thesis was to understand how IoT deployment in various settings can help in analysis of pollution data and provide valuable clues on its causative factors. The latter was achieved by doing both spatial neighbourhood and temporal analysis on these data-sets. The research showcases a comparative analysis between stationary and mobile IoT node deployment by carefully performing case studies of respective deployment strategies. In the study shown in chapter 3, temporal variation of PM concentration is analyzed in order to understand the PM data that is captured during the COVID-19 lockdown using two stationary IoT networks deployed in the Hyderabad city. The temporal variation analysis showed by means of yearly and seasonal trends that the lockdown caused a steep decline in human-centric activities such as vehicular traffic, thus resulting in a downfall in PM values. Although it also highlighted that not all decline could be accounted to the lockdown as the temperature and PM were negatively correlated during that time, which happened to be the onset of summer. Chapter 4 talks about a mobile IoT deployment case study where the focus is on the spatial variation of PM data. The analysis highlighted the impact of commercial/industrial activities, settlement patterns and vegetation profile around a location by evaluating the corresponding PM distribution.

The research was initially approached to study the two distinct types of IoT node deployment strategies and their respective distinction in terms of data collection and analysis. This led to the evaluation of factors that account for PM variation in respective modes of deployment. This work, by performing separate sets of studies, highlights the different ways by which ambient PM levels can be influenced. The importance of both spatial and temporal variation of PM captured by mobile and stationary IoT nodes respectively provides a holistic picture of PM levels of a location, the factors that influence it and approximately by how much. In most of the IoT-based air pollution research, mostly the temporal data is subjected to in-depth analysis and the spatial data is either studied in a limited way or completely overlooked. By bringing these together, it has helped in improving the understanding of causes.

There are a lot of lessons and inferences that the results presented in this thesis contributes. In stationary IoT deployment, it can be seen how historical data collected over a course of time enables to understand the PM behavior of a location. It also showed how much influence meteorological data such as temperature and humidity have over ambient PM levels. The mobile IoT deployment accounted for the influence of spatial and anthropogenic factors in deciding the PM variation of a location.

Although the work presented in this research yielded some insightful results in terms of explaining PM behavior, there are still a number of challenges that this work has not tackled successfully. There are a few generic challenges (common to both stationary and mobile IoT deployment) such as loss of data due to sensor failure and internet connectivity. Some specific challenges in stationary IoT deployment include devising strategies to ensure optimum spatio-temporal resolution of the stationary IoT network deployment and dealing with long-term maintenance issues of the network to ensure seamless collection of historical data. Mobile IoT deployment faces a wide range of challenges, some of which are, tackling bias in data caused due to inconsistency in vehicle speed where mobile node is mounted and deciding the best fit sampling strategy before the mobile node data collection campaign.

5.2 Future Scope

Although the research presented in this thesis appears exhaustive, it has barely scratched the surface. During the course of the research, lots of challenges have been encountered, thus providing abundant scope of work to be done in the days to come. Adoption of efficient deployment and sampling strategies while setting up a stationary IoT network for pollution monitoring in a given area based on the geospatial and anthropogenic features of the area is one of the biggest challenges that could be addressed. Future researchers can come up with best fit sampling strategy to handle dynamic data obtained from mobile IoT node, which can be extended as a generic strategy for moving object databases. Another possible scope would be to explore the role of factors like speed of the vehicle carrying the mobile IoT node in influencing the PM values and whether it creates a bias in the measuring the ambient PM levels compared to a stationary node. Other pollutants could be identified that could have direct or indirect influence over ambient PM levels. Lastly, coming up with a hybrid pollution monitoring system comprising of both stationary and mobile IoT nodes for optimum spatio-temporal resolution and pollution accountability.

Related Publications

Conference Papers:

- S. Deb, C. Rajashekar Reddy, S. Chaudhari, K. Vemuri, and K.S. Rajan, "IoT network based analysis of variations in particulate matter due to covid-19 lockdown", *IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2021.
- S. Deb, A. Dwivedi, S. Chaudhari, and K.S. Rajan, "Spatial Factor Analysis of Mobile IoT Data: A Case Study on PM Across India", *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022.

Bibliography

[1] *Internet of Things*, https://www.ibm.com/blogs/internet-of-things/what-is-the-iot/.

[2] K. Adams et al., "Particulate matter components, sources, and health: Systematic approaches to testing effects," *Journal of the Air and Waste Management Association*, pp. 544–558, 2015.

[3] I. Yadav and N. Devi, "Biomass burning, regional air quality, and climate change," in *Encyclopedia of Environmental Health (Second Edition)*, J. Nriagu, Ed. Oxford: Elsevier, 2019, pp. 386–391. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978012409548911022X

[4] A. S. Moursi et al., "An IoT enabled system for enhanced air quality monitoring and prediction on the edge," *Springer Complex Intelligent Systems*, pp. 2923–2947, 2021.

[5] S. J. Johnston et al., "City Scale Particulate Matter Monitoring Using LoRaWAN Based Air Quality IoT Devices," *MDPI Sensors*, vol. 19, no. 1, pp. 1424–8220, Jan 2019.

[6] C. R. Reddy et al., "Improving Spatio-Temporal Understanding of Particulate Matter using Low-Cost IoT Sensors," in *IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp. 1–7.

[7] *CPCB website*, accessed 1 July 2020, https://www.cpcb.gov.in/.

[8] R. Kamal, *Internet of Things: Architecture and Design Principles*. Mc Graw Hill Education, 2017.

[9] P. Lea, *Internet of Things for Architects*. Birmingham, UK: Packt Publishing Ltd, 2018.

[10] C. C. Sobin, "A Survey on Architecture, Protocols and Challenges in IoT," *Wireless Personal Communications*, vol. 112, 2020.

[11] D. Norris, *The Internet of Things*. Mc Graw Hill Education, 2015.

[12] A. Bahgya and V. Madisetti, *Internet of Things: A Hands-on Approach*. Universities Press, 2015.

[13] T. Kramp et al., *Introduction to the Internet of Things*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–10. [Online]. Available: https://doi.org/10.1007/978-3-642-40403-0_1

[14] "Internet Of Things (IoT)." [Online]. Available: https://www.gartner.com/en/information-technology/glossary/internet-of-things

[15] "Internet of Things Global Standards Initiative." [Online]. Available: https://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx

[16] *Major components of IoT*, accessed 2 May 2022, https://www.rfpage.com/what-are-the-major-components-of-internet-of-things/.

[17] T.-H. Kim, C. Ramos, and S. Mohammed, "Smart City and IoT," *Future Generation Computer Systems*, vol. 76, pp. 159–162, 2017.

[18] H. Arasteh et al., "IoT-based smart cities: A survey," in *IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, 2016, pp. 1–6.

[19] H. Rajab and T. Cinkelr, "IoT based Smart Cities," in *International Symposium on Networks, Computers and Communications (ISNCC)*, 2018, pp. 1–4.

[20] J. Markendahl, S. Lundberg, O. Kordas, and S. Movin, "On the role and potential of IoT in different industries: Analysis of actor cooperation and challenges for introduction of new technology," in *Internet of Things Business Models, Users, and Networks*, 2017, pp. 1–8.

[21] L. D. Xu, W. He, and S. Li, "Internet of Things in Industries: A Survey," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.

[22] "20 Examples of Wearables and IoT Disrupting Healthcare," Dec 2020. [Online]. Available: https://www.avenga.com/magazine/wearables-iot-healthcare/

[23] R. K. Kodali, G. Swamy, and B. Lakshmi, "An implementation of IoT for healthcare," in *IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2015, pp. 411–416.

[24] M. Haghi Kashani, M. Madanipour, M. Nikravan, P. Asghari, and E. Mahdipour, "A systematic review of IoT in healthcare: Applications, techniques, and trends," *Journal of Network and Computer Applications*, vol. 192, p. 103164, 2021.

[25] D. Metcalf, S. T. Milliard, M. Gomez, and M. Schwartz, "Wearables and the Internet of Things for Health: Wearable, Interconnected Devices Promise More Efficient and Comprehensive Health Care," *IEEE Pulse*, vol. 7, no. 5, pp. 35–39, 2016.

[26] F. John Dian, R. Vahidnia, and A. Rahmati, "Wearables and the Internet of Things (IoT), Applications, Opportunities, and Challenges: A Survey," *IEEE Access*, vol. 8, pp. 69 200–69 211, 2020.

[27] X.Krasniqi and E.Hajriz, "Use of IoT Technology to Drive the Automotive Industry from Connected to Full Autonomous Vehicles," *17th IFAC Conference on International Stability, Technology and Culture TECIS 2016*, vol. 49, no. 29, pp. 269–274, 2016.

[28] B. V. Philip, T. Alpcan, J. Jin, and M. Palaniswami, "Distributed Real-Time IoT for Autonomous Vehicles," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1131–1140, 2019.

[29] D. Minovski, C. Åhlund, and K. Mitra, "Modeling Quality of IoT Experience in Autonomous Vehicles," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 3833–3849, 2020.

[30] X. Chen, J. Liu, X. Li, L. Sun, and Y. Zhen, "Integration of IoT with smart grid," in *IET International Conference on Communication Technology and Application (ICCTA 2011)*, 2011, pp. 723–726.

[31] S. McGrath, C. Flanagan, L. Zeng, and C. O'Leary, "Iot personal air quality monitor," in *2020 31st Irish Signals and Systems Conference (ISSC)*, 2020, pp. 1–4.

[32] J. Jo, B. Jo, J. Kim, S. Kim, and W. Han, "Development of an iot-based indoor air quality monitoring platform," *Journal of Sensors*, vol. 2020, pp. 1–14, 01 2020.

[33] S. Dhingra et al., "Internet of Things Mobile–Air Pollution Monitoring System (IoT-Mobair)," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5577–5584, June 2019.

[34] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin, "An lstm-based aggregated model for air pollution forecasting," *Atmospheric Pollution Research*, vol. 11, 05 2020.

[35] T. Xayasouk, H. Lee, and G. Lee, "Air pollution prediction using long short-term memory (lstm) and deep autoencoder (dae) models," *Sustainability*, vol. 12, p. 2570, 03 2020.

[36] U. Mahalingam et al., "A machine learning model for air quality prediction for smart cities," in *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, 2019, pp. 452–457.

[37] A. Mbarak, Y. Yetis, and M. Jamshidi, "Data - based pollution forecasting via machine learning: Case of northwest texas," in *2018 World Automation Congress (WAC)*, 2018, pp. 1–6.

[38] J. Lee et al., "Empirical analysis of tree-based models for pm¡inf¿2.5¡/inf¿ concentration prediction," in *2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2019, pp. 1–7.

[39] X. Xu and M. Yoneda, "Wrf model based quantized kernel adaptive filter for pm¡inf¿2.5¡/inf¿ concentration online prediction," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2019, pp. 994–1001.

[40] V.-D. Le, T.-C. Bui, and S.-K. Cha, "Spatiotemporal deep learning model for citywide air pollution interpolation and prediction," in *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2020, pp. 55–62.

[41] J. Van den Bossche et al., "Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset," *Atmospheric Environment*, vol. 105, pp. 148–161, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231015000254

[42] I. Wadlow et al., "Understanding spatial variability of air quality in sydney: Part 2 – A Roadside Case Study," *Atmosphere*, vol. 10, no. 4, 2019. [Online]. Available: https://www.mdpi.com/2073-4433/10/4/217

[43] X. Wu et al., " Exposure to air pollution and COVID-19 mortality in the United States," *medRxiv*, Apr 2020.

[44] H. Guo, S. Kota, S. Sahu, and H. Zhang, "Contributions of local and regional sources to pm2.5 and its health effects in north india," *Atmospheric Environment*, vol. 214, p. 116867, 2019.

[45] I. Kheirbek et al., "The contribution of motor vehicle emissions to ambient fine particulate matter public health impacts in New York City: a health burden assessment," *Environ. Health*, 2016.

[46] S. Mahato et al., "Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India," *Elsevier*, Apr. 2020.

[47] J. Berman et al., "Changes in u.s. air pollution during the covid-19 pandemic," *Science of The Total Environment*, vol. 739, p. 139864, 2020.

[48] L. Menut et al., "Impact of lockdown measures to combat covid-19 on air quality over western europe," *Science of The Total Environment*, vol. 741, p. 140426, 2020.

[49] *Google Maps*, accessed 13 Sep. 2020, https://www.google.co.in/maps.

[50] *CPCB node sepcifications*, http://uppcb.com/pdf/volume_2_20nov14.pdf.

[51] *SDS011 Nova Sensor Specifications*, accessed 06 Aug. 2020, http://www.inovafitness.com/en/a/chanpinzhongxin/95.html.

[52] *ThingSpeak*, accessed 06 Aug. 2020, https://thingspeak.com/.

[53] C. Salgado, C. Azevedo, H. Proença, and S. Vieira, *Noise Versus Outliers*. Cham: Springer International Publishing, 2016, pp. 163–183.

[54] Z. Lu and K.-H. Yuan, *Welch's t test*, 01 2010, pp. 1620–1623.

[55] *Hyderabad making most of lockdown to build roads, flyovers*, 15 Apr. 2020, accessed 6 Jul. 2020, https://www.livemint.com/news/india/hyderabad-making-most-of-lockdown-to-build-roads-flyovers-11586959814540.html.

[56] S. Gautam and J. Brema, "Spatio-temporal variation in the concentration of atmospheric particulate matter: A study in fourth largest urban agglomeration in india," *Environmental Technology Innovation*, vol. 17, p. 100546, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352186419304262

[57] S. Somvanshi and M. Kumari, "Comparative analysis of different vegetation indices with respect to atmospheric particulate pollution using sentinel data," *Applied Computing and Geosciences*, vol. 7, p. 100032, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590197420300148

[58] *NodeMCU Datasheet*, https://www.espressif.com/sites/default/files/documentation/0a-esp8266ex_datasheet_en.pdf.

[59] *SDS011 Nova PM Sensor*, https://cdn-reichelt.de/documents/datenblatt/X200/SDS011-DATASHEET.pdf.

[60] *Neo6 GPS Sensor*, https://www.u-blox.com/sites/default/files/products/documents/NEO-6_DataSheet_(GPS.G6-HW-09005).pdf.

[61] *NDVI*, https://www.usgs.gov/landsat-missions/landsat-normalized-difference-vegetation-index.

[62] *USGS Earth Explorer*, 15 Apr. 2020, accessed 6 Jul. 2021, https://earthexplorer.usgs.gov/.

[63] *Bharatmala Project*, accessed 4 Sep. 2021, https://morth.nic.in/sites/default/files/PragatiKiNayiGati/pdf/maharashtra.pdf.