JADAVPUR UNIVERSITY

MASTER DEGREE THESIS

Real-time Participatory Sensing based Noise Pollution Monitoring system for Smart City

A thesis submitted in fulfillment of the requirements for the degree of Master of Engineering

in

Computer Science and Engineering

by

Abhimanyu Kumar
Exam Roll No.: M4CSE19025
Class Roll No.: 001710502021
Registration No.: 140760 of 2017-18

Under the Guidance of

Dr. SARBANI ROY
Associate Professor
Department of Computer Science and Engineering
Jadavpur University

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Kolkata - 700032

May 21, 2019

Declaration of Authorship

I, Abhimanyu Kumar, declare that this thesis titled, "Real-time Participatory Sensing based Noise Pollution Monitoring system for Smart City" and the work presented in it are of my own. I confirm that:

- This work was done wholly or mainly while in candidature for a master degree in computer science and engineering at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entire of my own work.
- I have acknowledged all the main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

To Whom It May Concern

This is to certify that the thesis entitled "Real-time Participatory Sensing based Noise Pollution Monitoring system for Smart City" is a bona-fide record of work carried out by Abhimanyu Kumar, Examination Roll No.: M4CSE19025, University Registration No.: 140760 of 2017-18 in partial fulfilment of the requirements for the award of the degree of Master of Engineering in Computer Science and Engineering from the Department of Computer Science and Engineering, Jadavpur University for the academic session 2017-2019. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

Dr. Sarbani Roy
Associate Professor
Department of Computer Science and Engineering
Jadavpur University
Kolkata - 700032

Prof. Mahantapas Kundu
Head of the Department
Department of Computer Science and Engineering
Jadavpur University
Kolkata - 700032

Prof. Chiranjib Bhattacharjee
Dean, Faculty of Engineering and Technology
Jadavpur University
Kolkata - 700032

Certificate of Approval

(Only in case the thesis is approved)

The thesis at instance is hereby approved as a creditable study of an engineering subject carried out and presented in a manner satisfactory for its acceptance as a perquisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve this thesis only for the purpose for which it is submitted.

____

(Sign of Examiner)
Date:

____

(Sign of Examiner)
Date:

"Take risks in your life. If you win, you can lead. If you lose, you can guide!"

Swami Vivekananda

JADAVPUR UNIVERSITY

Abstract

Faculty of Engineering and Technology
Department of Computer Science and Engineering

Master of Engineering

Real-time Participatory Sensing based Noise Pollution Monitoring system for Smart City

by Abhimanyu Kumar

The behaviour of natural phenomena is related to spatial distributions and temporal conditions and results in spatiotemporal datasets. Nowadays, dealing with spatiotemporal datasets has become an important criterion. We can use spatial interpolation techniques to estimate natural phenomena. Using spatial interpolation techniques, we can develop noise pollution monitoring applications. In this work, our main aim is to generate a participatory sensing based real-time noise pollution map of the urban areas in an efficient way. The noise map generation is an effective way for visualizing and accessing the noise pollution level of the urban areas. In this thesis, we have proposed three different spatial interpolation methods such as Grid-MSM, I-Grid-MSM, and Grid-MSM-SK. These methods work on the participatory based framework, whose task is to generate a real-time dynamic noise pollution map of the urban areas. We evaluate these methods on a real-time participatory sensing based noise datasets which is collected by participants (volunteers and citizen) over a period of one year near Jadavpur University, Kolkata, India. The results of these methods are compared with baseline methods like inverse distance weighting (IDW) and Simple Kriging (SK). The result shows that on average I-Grid-MSM is a suitable method for inferring noise pollution map for the urban areas. I-Grid-MSM is a feasible and low-cost noise inferring solution.

Acknowledgements

I would like to extend my heartfelt gratitude to people who helped to bring this thesis work to complete. First, I would express my deep and sincere gratitude and appreciation to my supervisor Dr. Sarbani Roy without whose continuous support and encouragement this work would not have been possible. Her assistance, valuable suggestions and personal guidance throughout the duration of the project has played a pivotal role, without which I would never have been able to reach this far.

I would like to express my gratitude to Mr. Joy Dutta, Visvesvaraya Research Fellow, Jadavpur University, Kolkata and my senior Mr. Asif Iqbal Middya, Ph.D. scholar, Jadavpur University, Kolkata and also thankful to my senior Mr. Subhaisish Ghosh, Master of Engineering, Jadavpur University for brainstorming with me in this work.

I would also wish to thank Prof. Mahantapas Kundu, Head of the Department of Computer Science and Engineering, Jadavpur University and Prof. Chiranjib Bhattacharjee, Dean, Faculty of Engineering and Technology, Jadavpur University for providing me all the facilities and for their support to the activities of this research.

During the last year I had a great pleasure to work in a DST-FIST laboratory and I am grateful to all the members of the DST-FIST laboratory for their kind co-operation and help.

I am thankful to my parents and my brother who have always been my constant source of support and inspiration.

Last, but not least, I would like to thank all my friends, classmates and all respected teachers for their valuable suggestions and helpful discussions.

Regards,
Abhimanyu Kumar
Exam Roll No.: M4CSE19025
University Registration No.: 140760 of 2017-18
Department of Computer Science and Engineering
Jadavpur University

Contents

List of Figures

List of Tables

List of Abbreviations

CCTV	**C**losed-**C**ircuit **T**ele**v**ision
CPCB	**C**entral **P**ollution **C**ontrol **B**oard
EEA	**E**uropean **E**nvironment **A**gency
Grid-MSM	**G**rid based **M**odified **S**hepard's **M**ethod
Grid-MSM-SK	**G**rid based **M**odified **S**hepard's **M**ethod-**S**imple **Kriging**
IDW	**I**nverse **D**istance **W**eighted
I-Grid-MSM	**I**terative-**G**rid based **M**odified **S**hepard's **M**ethod
MGRS	**M**ilitary **G**rid **R**eference **S**ystem
NANMN	**N**ational **A**mbient**N**oise **M**onitoring **N**etwork
NIHL	**N**oise **I**nduced **H**earing **L**oss
NIPTS	**N**oise-**I**nduced **P**ermanent **T**hreshold **S**hift)
NMS	**N**oise **M**onitoring **S**ystem
NYC	**N**ew **Y**ork **C**ity
RMAE	**R**elative **M**ean **A**bsolute **E**rror
RMSE	**R**oot **M**ean **S**quare **E**rror
SK	**S**imple **K**riging
SLM	**S**ound **L**evel **M**eter
SPL	**S**ound **P**ressure **L**evel
TRAI	**T**elecom **R**egulatory **A**uthority of **I**ndia
UPS	**U**niversal **P**olar **S**tereographic
UTM	**U**niversal **T**ransverse **M**ercator
WHO	**W**orld **H**ealth **O**rganization

Physical Constants

Reference sound pressure level $p_O =$ typically 20μ Pascal

List of Symbols

S	Sample data.
$\Re$	Radius within neighbour is considered for computing weight.
NB_{Th}	The threshold value for number of nearest neighbour.
D	The dimension of each grid.
G	Grids.
NN_i	Set of all nearest neighbour for i^{th} grid within radius R.
$d(mid_i, mid_j)$	distance between midpoint of i^{th} grid and midpoint of j^{th} grid.
w_{ij}	the weight of j^{th} neighbour from i^{th} grid.
GN_{pred_i}	Predicted noise value for i_{th} Grids.
g_i	i^{th} grid.
S_{ij}	j^{th} neighbour of i^{th} grid.
mid_i	Midpoint of i^{th} grid.
CBD	Chess Board Distance
$Grid_{Th}$	Threshold value for I-Grid-MSM
C	A counter variable for I-Grid-MSM
$\widehat{\gamma}(h)$	The estimated semivariance at a separation distance h.
$models$	Modelling Function(spherical, exponential, gaussian).
hs	kriging distance.
bw	kriging bandwidth.
K	Number of neighbour points to consider for krig.
m	Represent a mean value
$\mathbf{w_1}$	A covariance matrix using model.
$\mathbf{w_2}$	A covariance matrix between data point and unpredicted midpoint g_i .
λ	A weighting factor for krig

I dedicate this to my parents for their continuous support throughout my journey!

1 INTRODUCTION

1.1 Background

The monotonic increase in the growth of population leads to the expansion of urban cities and industries which not only contemporize people's live, but it also brings some blight. The rapid growth of urbanization has driven various health issues and also affect the environment and animal habitats. During the past decade (2001-2011), the growth of the urban population in India increased by 31.8% [1]. Such population growth has led to increasing transportation demands, vehicular increase, congestion of roads, machinery productions and increment in home appliances due to this it led to the different types of environmental pollutions which have an adverse effect on citizen's life. In today's world scenarios, noise pollution is one of the major problems in urban environments. Under the air (Prevention and Control of Pollution) Act, 1981 [2] noise is considered an air pollutant. The word "noise" descends from the Latin word "nausea," meaning seasickness, or, more generally, any similar sensation of disgust, annoyance, or discomfort. According to CPCB (Central Pollution Control Board), "Noise is usually defined as an unwanted sound pollutant which produces undesirable physiological and psychological effects in an individual, by interfering with one's social activities like work, rest, recreation, sleep, etc.". A sound might be unwanted because it is loud, unpleasant, annoying, intrusive, etc. Noise pollution is nothing but undesired environmental sounds created by activities of human and human-made machines or devices, which may cause to various serious health problems like hearing loss or impairedness, increasing stress levels, behavioral and mental problems, insomnia, heart ailments, hypertension, and many more [3] [4]. It also affects the behaviour and habitat of animals.

According to WHO (World Health Organization) guidelines, it is recommended to use community noise below 30 dB[A] (A-weighted decibels) for a good sleep in a bedroom during the night and in the classroom, it should be less than 35 dB[A] for good teaching. The annual average late night noise outside the bedrooms should be less than 40 dB[A] [5]. In the recent report of WHO, noise is considered as one of the prime environmental hazards to both physical and mental health and well-being [6]. The CPCB guideline for India is shown in table 1.1. According to a 2011 WHO estimate, almost 6% of people in India suffer from hearing loss. Prolonged exposure to noise above 60 dB can lead to irreversible NIHL (Noise Induced Hearing Loss). The report of EEA (European Environment Agency) clearly unveils that every year more than 16600 premature death in Europe is due to the environmental noise. Approximately 13 million adults are suffering from sleep disturbance and 32 million

TABLE 1.1: Ambient air quality standards in respect of noise

Area Code	*Category of Area/Zone*	**Limit in dB(A) Leq***	
		Day Time	*Night Time*
A	Industrial Area	75	70
B	Commercial Area	65	55
C	Residential Area	55	45
D	Silence Zone	50	40

*dB(A) Leq denotes the time weighted average of the level of sound in decibels on scale A which is relatable to human hearing.
A "decibel" is a unit in which noise is measured. "A", in dB (A) Leq, denotes the frequency weighting in the measurement of noise and corresponds to frequency response characteristics of the human ear.
Leq: It is energy mean of the noise level over a specific period. Source: Central Pollution Control Board, India

annoyed by it. Almost 15% population (26 million people) of U.S.A may have suffered from NIPTS (noise-induced permanent threshold shift) which is a permanent reduction in hearing acuity due to excessive exposure to workplace or leisure noise [7].

1.2 Motivation

In today's world, noise pollution becomes one hazardous environmental problem. Most of the citizen are unaware from adverse effect caused by noise pollution and they are exposed to an environment having a high level of noise pollution, which can cause different types of a long term or short term serious health hazard and also degrades the quality of urban lifestyle. We need to monitor noise pollution and generate noise pollution maps especially in urban and industrial areas to aware citizens from aforesaid problems. But, it is a difficult task to completely and accurately monitor urban noise pollution in a large region. There may be different ways through which we can monitor noise pollution in urban areas. One way is the traditional approach in which few noise monitoring stations were situated in specific locations (e.g. near crowded streets, railway stations, hospitals, airports, interesting residential and commercial areas) in the urban area. Data are collected for a prolonged time (few days/month/year) using these monitoring stations and using some mapping techniques, they generate noise maps by extrapolating local noise data to far-ranging areas. One thing is noticeable that these monitoring stations can achieve a high level of precision but only for the local perspective. Along with this there are a lot of demerits regarding data collection strategies of the traditional approach and there are the following:

- Data is collected at very sparse geographical locations which are not a proper way to generate noise pollution maps of high spatial and temporal resolution.
- These kinds of data collection strategies only cover outdoor noise but most of the people spend their time indoors. Sometimes indoor noise becomes more annoying and unbearable and should be considered.
- The context (mainly temporal) in the urban region changes rapidly if the contextual location is far away from monitoring stations it will become very difficult to monitor such local or short-duration noise effects.
- The overall cost for monitoring noise pollution using traditional methods is very expensive due to the need for human expertise, expensive equipment,

periodic maintenance of equipment and effort required to monitor noise pollution.

Another way of collecting noise data is simulations based on inputs such as traffic-flow data, road or rail type, vehicle type, machinery type and many more. This approach is significantly useful because it allows the assessment of background noise in the absence of physical data with adequate granularity. However, this approach suffers from some problems, one of the main problem these type of data sources in urban cities is limited and because of its simulations, this approach does not consider any kind of incidental sound caused by traffic jams, construction work, whirring sound, the assembled crowd, and other unexpected sounds. This approach does not include indoor noise and most of the region which should be considered regarding noise data collection.

To conquer aforesaid limitations, we can use a new paradigm of sensing called participatory sensing. Participatory sensing is a new pattern of sensing in which citizen(s) and communities use their smart devices (smartphones, tablets, etc.) to collect data and cloud services to store, integrate and analyze systematic data and their pattern to explore any phenomena of interest [8]. Nowadays, we all are surrounded by smart devices and these devices become an essential part of our daily lives. Smart devices like smartphones and tablets equipped with various sensors like GPS, accelerometer, gyroscope, magnetometer, proximity, etc. Due to the availability of such smart devices to nearly every person makes an easy task for collecting data in a remote location within the city. So using these smart devices we can achieve a huge amount of data which will definitely solve the problem of sparsity. According to TRAI (Telecom Regulatory Authority of India) report 31st March 2019, approx. 1181.97 million wireless telephone subscribers in India, which is the second largest in the world after China. The report of TRAI is shown in figure 1.1. The overall teledensity (circle/state wise) is shown in figure 1.2. The teledensity of Delhi service area is recorded as the maximum which is 237.57 % at the end of January-2019 while the Bihar service area has a minimum teledensity of 62.54 during the same period [9]. The urban wireless teledensity is more than as compared to that of the rural area shown in figure 1.3. These reports clearly show that there are a large number of smartphone users in urban areas and using the concept of participatory sensing we can collect context-aware noise data in the urban area, which definitely discloses the problem of sparsity and local noise effect. For the analysis of such huge amount of data, we can also use spatiotemporal, machine learning, deep learning, and context-based methodologies so that will accurately and completely infer the noise map of urban areas to aware citizen from adverse effect caused by noise pollution. This concept will definitely improve the lifestyle of the urban.

1.3 Objectives

To address the aforesaid problem regarding traditional methods we have proposed our system. The main objective of our proposed system is precisely described below:

- Under the flag of a participatory system, noise data is collected using several smartphones or tablets at a very precise level so that the issue of data sparsity will be minimized.
- Collect and store noise data in the cloud to analyze systematic behaviour of noise data and kept those records for future perspectives.

Particulars	Wireless	Wireline	Total (Wireless+ Wireline)
Total Telephone Subscribers (Million)	**1181.97**	**21.79**	**1203.77**
Net Addition in January, 2019 (Million)	5.97	-0.07	5.90
Monthly Growth Rate	0.51%	-0.34%	0.49%
Urban Telephone Subscribers (Million)	**654.20**	**18.71**	**672.91**
Net Addition in January, 2019 (Million)	6.68	-0.05	6.63
Monthly Growth Rate	1.03%	-0.24%	1.00%
Rural Telephone Subscribers (Million)	**527.77**	**3.08**	**530.86**
Net Addition in January, 2019 (Million)	-0.70	-0.03	-0.73
Monthly Growth Rate	-0.13%	-0.93%	-0.14%
Overall Tele-density*(%)	**90.15**	**1.66**	**91.82**
Urban Tele-density*(%)	156.85	4.49	161.34
Rural Tele-density*(%)	59.04	0.34	59.38
Share of Urban Subscribers	55.35%	85.86%	55.90%
Share of Rural Subscribers	44.65%	14.14%	44.10%
Broadband Subscribers (Million)	**521.77**	**18.27**	**540.04**

FIGURE 1.1: Highlights of Telecom Subscription Data as on 31st January, 2019. Source: TRAI

- Generate noise pollution maps at a very precise level by using spatial interpolation techniques.
- Infer noise pollution for those areas where data is not collected.
- Generate accurately and completely noise pollution map of the city based on past activities at low cost-effectively and efficiently in near real time.
- Aware citizen(s) regarding noise pollution so that they can visualize their own and others exposed to the environment. We can also recommend them so that they can make a better decision.

1.4 Contributions

The main contribution of this work is to design a participatory sensing based framework in which we can apply some spatial interpolation methodologies in order to achieve an accurate and completely noise pollution map of the city. In this thesis, we have proposed a different variety of methodologies like Grid-MSM, I-Grid-MSM, and Grid-MSM-SK. Instead of proposing one method, we have proposed a different verity of methods, it is because each method has its own benefit (see section 5.4). For example, when a sufficient amount of sample data lies within the range of interpolating points (in our case midpoint of grids) then we can easily apply Grid-MSM otherwise I-Grid-MSM can be applied (see section 5.4). We divide the urban region into a smaller grid (of equal size) for the purpose of noise map generation. It seems very impractical to predict noise pollution at every GPS location. Therefore, we use MGRS (Military Grid Reference System) coordinates system to convert geo (latitude and longitude) coordinates into MGRS coordinates (grid designator, easting, and northing) to split the area into the small grid (see section 4.3). As noise is inversely proportional to distance, so we can use the spatial interpolation technique. We can also use machine learning (future work) based methods for learning

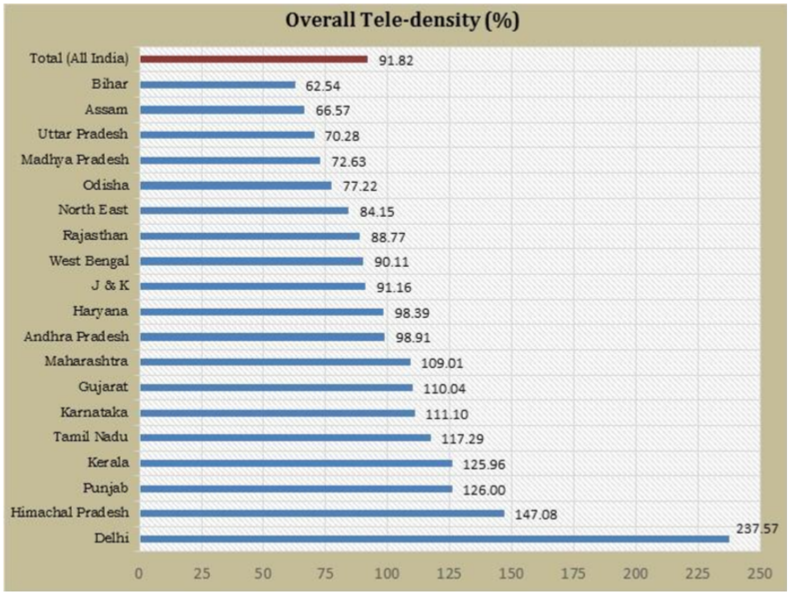

FIGURE 1.2: Overall Tele-density (Circle/State Wise) – As on 31st January, 2019. Source: TRAI

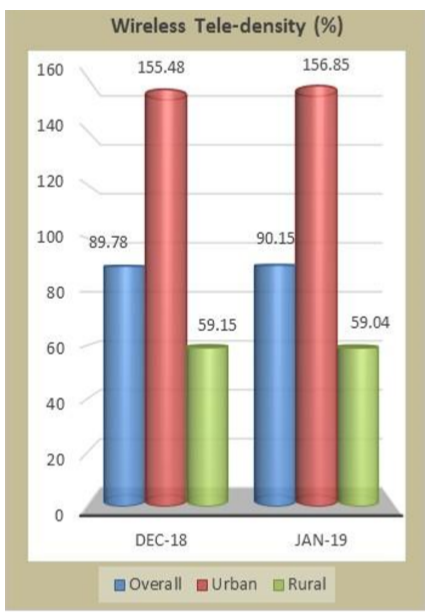

FIGURE 1.3: Wireless Tele-density urban Vs rural. Source: TRAI

and predicting/inferring noise of grids. Apart from this, the noise pollution in urban is time and context dependent, so including temporal and contextual features in the aforesaid solution will definitely improve noise inferring accuracy. In the spatial method, we first use Grid-MSM (Grid-based Modified Shepard's Method) to predict noise value of those grids' having a sufficient number of noise samples within some specific range, but this method suffers from some limitations (see section 5.4.1). To overcome Grid-MSM we use I-Grid-MSM and Grid-MSM-SK. I-Grid-MSM method iteratively propagates its neighbour values which help to infer noise for unknown grids. This process continues until it reaches some threshold value. While Grid-MSM-SK uses the concept of the kriging technique. Moreover, these methods have their own benefits depending upon the dataset and feature value. The main idea of this thesis is to dynamically generate a noise pollution map of the city based on real-time participatory sensing at very low cost in an efficient way with high accuracy. Our key contribution is precisely described below:

- We have proposed participatory sensing based spatial interpolation methodologies (Grid-MSM, I-Grid-MSM, and Grid-MSM-SK) for generating dynamic noise maps.
- We also design a participatory based framework using the concept discussed in [12], which provides users to get the noise map as a service based on their query.
- I-Grid-MSM is most the suitable approach for this purpose.

1.5 Organization of the Thesis

The rest of the thesis is organized as follows: Chapter 2 consists of a literature survey corresponding to noise pollution monitoring strategies, some key potential of participatory sensing, and some existing participatory sensing based spatial interpolation methods. Spatial interpolation techniques are discussed in chapter 3. In chapter 4, participatory sensing based framework, data collection, cleaning strategies, and our proposed methodologies are discussed. In Chapter 5, we describe the details of implementation, noise pollution map generation and results from the comparison. Finally, in chapter 6, we conclude our work and some outlook on future work.

2 Literature Survey

Recall from chapter 1, noise pollution becomes one of the top environmental pollution hazards in urban. So it mandatory to monitor noise pollution in the urban areas. There are several ways to monitor noise pollution like the traditional approach and participatory sensing based approach. Few countries are still using the traditional approaches for monitoring noise in urban and some of them are using a participatory-based noise pollution strategy. But, the main question is how to monitor noise pollution accurately and completely when the region is large enough. Here we provide some literature survey regarding different approaches to noise pollution monitoring.

2.1 Traditional Method for noise pollution monitoring

In the case of the traditional method for noise pollution monitoring, few noise monitoring stations are situated in some specific locations such as near crowded streets, railway stations, hospitals, airports, interesting residential and commercial areas. They collect noise data for a prolonged time which will include a few days, a few months or in some cases a few years. Based on these collected data they provide an urban noise map by using some kind of interpolation or extrapolation technique. In India, the Central Pollution Control Board (CPCB) implements and successfully manages the National Ambient Noise Monitoring Network (NANMN) program in 2016 and 2017. They have used a traditional approach with some modifications. Their main object is to monitor ambient noise pollution over Indian and to find collective significant factor and its collective impacts as to help the concern authorities and planners for decision making at the pro-active stage. For this purpose, they install noise monitoring stations at some selected locations. NANMN comprises 70 noise monitoring stations, which includes 7 metropolitan cities (Bangalore, Chennai, Delhi, Hyderabad, Kolkata, Lucknow and Mumbai) having 10 stations in each city. Figure 2.1 shows the satellite map of the monitoring station in Kolkata. The stations were installed as per the categorization of Ambient Noise standards (Silence, Residential, Commercial and Industrial) [1].

They have used traditional methodologies with advanced devices and technologies. They have used Sound Level Meter (SLM) which consists of a microphone, a processing unit, and real-out unit. They suggest that condenser microphone is most suitable for SLM which combines precision with stability and reliability. For measuring a large number of monitoring stations they came with a system called

Sl. No.	Station location	Category	Latitude	Longitude
1.	SSKM Hospital	Silence Zone	22°32' 19.58" N	88°20' 35.29" E
2.	Gole Park	Industrial Zone	22°31' 1.2" N	88°24' 15.8" E
3.	Kolkata HQ	Commercial Zone	22°33' 42.67" N	88°24' 32.46" E
4.	Patauli	Residential Zone	22°28' 21.07" N	88°23' 29.71" E
5.	New Market	Commercial Zone	22°33' 41.4" N	88°21' 10.4" E
6.	Birati N.	Residential Zone	22°40' 13.99" N	88°26' 1.74" E
7.	RG Kaur	Silence Zone	22°36' 16.18" N	88°22' 43.20" E
8.	Tollygunge	Commercial Zone	22°29' 56.48" N	88°20' 43.79" E
9.	Bag Bazar	Residential Zone	22°36' 4.61" N	88°22' 1.01" E
10.	Tartala	Industrial Zone	22°30' 56" N	88°18' 19.2" E

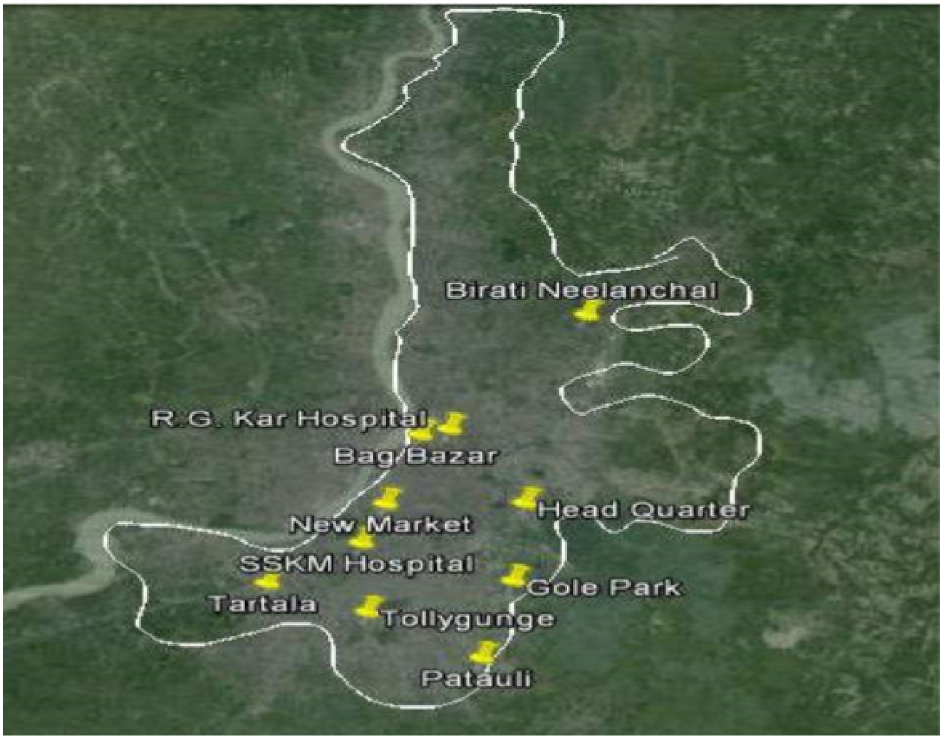

FIGURE 2.1: A satellite map of Noise monitoring stations of Kolkata and their corresponding GPS locations. Source: CPCB

Noise Monitoring System (NMS). This system is used for measuring real-time ambient noise. NMSs are mainly optimized for outdoor use which consists of a weatherproof cabinet containing a noise level analyzer and a battery, a communication device for transmitting data to receiving station, a back plate and an outdoor microphone (for measuring sound) all of which can be mounted on a mast. The high-quality microphone is connected to an advanced acquisition signal processing unit, completed with electronic measurement and processed-data storage unit, provided as well with an integrated GPRS modem [11].

In [11], results and observations are compiled and processed for interpretation on a daily, monthly and yearly basis. The results for Kolkata is shown in figure 2.2. This approach does not include any kind of interpolation or extrapolation technique. Considering only 10 locations in the urban does not provide a better noise pollution map.

	BAG BAZAR (R)		BIRATI (R)		GOLE PARK (I)		HEAD QUARTER (C)		NEW MARKET (C)	
	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time
Standard	**55**	**45**	**55**	**45**	**75**	**70**	**65**	**55**	**65**	**55**
No of Exceedance	362	363	346	364	13	120	24	365	345	336
No of Observations	362	363	364	364	345	342	365	365	345	336
Percentage of Exceedance	100	100	95	100	4	35	7	100	100	100
Max	79	77	76	70	79	77	68	63	84	82
Min	69	65	53	49	46	48	58	56	67	66
Median	75	69	60	54	69	69	63	57	73	73
L10	76	70	67	62	73	73	65	59	77	77
L50	75	69	60	54	69	69	63	57	73	73
L90	73	68	56	51	61	61	61	57	70	69

	PATAULI (R)		RGKAUR (S)		SSKM (S)		TARATALA (I)		TOLLYGUNJ (C)	
	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time
Standard	**55**	**45**	**50**	**40**	**50**	**40**	**75**	**70**	**65**	**55**
No of Exceedance	349	363	365	365	360	365	0	0	362	365
No of Observations	365	364	365	365	360	365	364	364	365	365
Percentage of Exceedance	96	100	100	100	100	100	0	0	99	100
Max	79	77	70	68	75	72	73	70	69	67
Min	48	42	61	57	61	55	59	57	64	61
Median	64	63	64	60	66	59	65	61	67	64
L10	73	72	66	62	69	63	68	64	68	65
L50	64	63	64	60	66	59	65	61	67	64
L90	59	57	63	58	63	57	62	60	66	63

FIGURE 2.2: Statistical behavior of sound level data of all 10 stations of Kolkata for the year 2017. Source: CPCB

2.2 Participatory Sensing

The word "Participatory" means to allow people, group, community to take part in or involved in an activity and "sensing" refers to perceive by sense or senses. The activity may be any kind of government program or some social issues like pollution control, health campaign, and educational encouragement, etc. The sensing activity can be done by using any kind of sensors for e.g., sensing noise level of a place we can use SLM (Sound Level Meter) device, finding some activities on social media or sensing some recent activities on social media we can use Facebook, Twitter, LinkedIn. Participatory sensing is a new pattern of sense in which people, groups and communities may use their smart devices for collecting data and these data are then stored in a cloud to analyze the systematic pattern, discover knowledge by applying some methodologies on data sets to generate new ideas. The general application model of participatory sensing is shown in figure 2.3. Using the potential of participatory sensing we can achieve some useful features regarding our society. Participatory sensing just not an idea it is a new paradigm sensing, which will definitely help for sensing data in any kind of platform. It allows individuals to measure, respond to and visualize their own and others exposure to environmental pollution (e.g. noise and components of air pollution) in near real time. Some key potential of participatory sensing is described below:

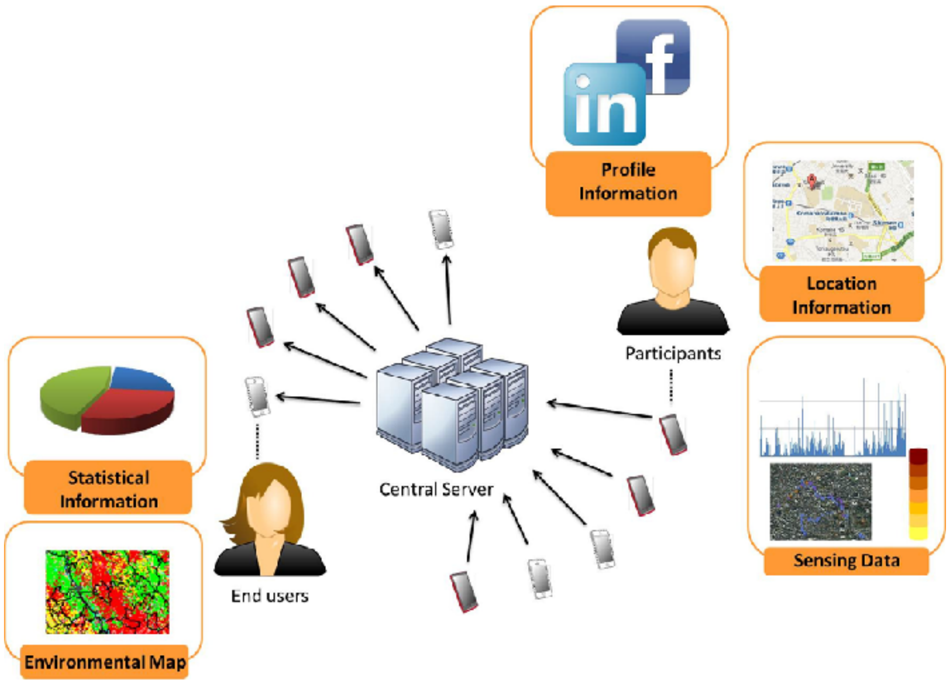

FIGURE 2.3: General application model of participatory sensing. Available from: https://www.researchgate.net/figure/General-application-model-of-participatory-sensing_fig1_272210676

2.2.1 New emerging sensing phenomenal for social and environment

In this section, we only focusing on participatory sensing related new sensing technologies and their technological possibilities for societies (urban) and its environment. The term sense related to the sensor. According to the Oxford English Dictionary, "a sensor is a device that detects or measures a physical property and records, indicates, or otherwise responds to it". In [20], they use sensing to refer to detecting signals that contain information on the physical, natural and social environment. So this definition includes two phenomena, first is physicochemical signals like sound, radiation, and images, etc., and the second is psychosocial signals such as behavior, opinions or moods, etc. They define sensing as collecting observations relating both to facts (objective) and to interpretations, opinions and moods (subjective) with the help of a sensor. Thus, they extend the definition of a sensor to include technological sensors as well as the human sensor.

The aforesaid definition of a technological sensor can be referred to a device that will capture physicochemical signals and then interpret them to a meaningful value. For example, a microphone translating a pressure into a noise level or a gas sensor translating a change in resistance into a gas concentration. These days, the huge development of technologies in data communication, it's storage and computing power allows us to generate and analyze the stream of tags, comments, votes, rating, opinion or counts. Thus we can define human sensors as the tools used to collect and extract meaning and information on the physical, natural or social environment from information registered by humans through text, numerical values or categories without the use of technological sensors. The different sections of this chapter will provide an outlook in different sensing domains and sensing technologies, and illustrate their potential. They deal with sensing systems that can be used actively (possibly with some training) by the general public, or to which the general public contributes by leaving their data (knowingly or unknowingly), and as such create new opportunities for collecting novel data, improving monitoring, and understanding the environment, human behaviour, opinions and moods.

Human sensor on the move

The human sensor on the move is nothing but mobile systems which are designed to collect data from smart devices (such as Smartphones or tablets) that users carry in their everyday life. Using these devices we can collect rich data about people's behavior and habits, as well as their environment. The rapid movement of people and devices in an urban environment brings them together into contact with each other and this massive movement generates rich information about urban. There are a lot of projects that have focused on this type of scenario. For example, the Reality Mining project [21] collected proximity, location and activity information, with proximity nodes being discovered through periodic Bluetooth scans and location information by cell tower IDs.

We can use smartphones as information gatekeepers because of its availability and usability, through which we can collect useful data. But we have to calibration our smartphones because it contain a cheap and low quality of sensors. For example, a smartphone microphone is meant to pick up voice, not to measure sound pressure. There is no wind cancellation and some frequency is suppressed, while others are enhanced. So calibration is an important task deal with a smartphone. Apart from calibration, energy consumption is also one the major factor. As we use a number of sensors at a time, become a major source of power consumption in modern smartphones and degrades the quality of devices. But, overall we can use a smartphone as the instrumentation of capturing data on move.

Environmental sensing

In this section, we only focus on environmental sensing and their approaches. Environmental sensing is sensing the behavior and sensing the pollution caused by human and human-made appliances or equipment. As the pollution level of urban grew day by day. Due to this, it led to the development of large scale monitoring networks to understand the sources, context, and dispersion of various kinds of pollution. Such a network improves environmental policy and helps in identifying the major source of pollution and their cause. The official monitoring networks are highly standardized using high-quality precision instruments. They are typically focused on a limited number of points of interest at which measurements are taken place with a high level of accuracy. This approach is well for long-term temporally averaged monitoring if the pollution concentration is only slightly influenced by local pollution sources. However, this approach is not well fitted for capturing spatial variability and short term fluctuations caused by local sources. However, we can overcome these challenges by using a monitoring network created with the help of miniaturized sensors that can be used as stand-alone devices, connected to smartphones or even embedded in smartphones. This concept of miniaturized sensors networks can be developed very cheaply and easily and hold the promise of enabling new kinds of intelligent networks called participatory sensing that allow monitoring of environmental parameters at significantly higher levels of spatiotemporal detail. For measuring the variability of spatiotemporal we can use a grid-based system based on user's requirements. Using the aforesaid concept we can monitor ambient air quality and also monitor noise pollution in urban areas. This is a huge potential of participatory based sensing through which we can sense the environmental pollutions.

Activity sensing of human being

This section basically includes the various observations of human activities through smart sensors that are connected to each other either directly or indirectly. For example, to monitor and guide the traffic activities we can use CCTV (Closed-circuit television) cameras or traffic signals, to monitor human activities that can use smart devices like Zephyr, HXL, etc., for surveillance and security purpose we use CCTV. We can also monitor traffic with the help of Floating car data. When observing humans outside vehicles, in multimodal transport or in indoor situations, scanning based on Bluetooth, WiFi, RFID tags or smartphones all offer a possible solution. So using Smart devices like CCTV, WiFi, Bluetooth, and many more, we can sense the activity of citizens in urban in order to monitor and guide traffic, provide better security, improve the quality of lifestyle.

Human sensors

This section will provide an overview and multiple cases of a system where humans are the primary source of information and themselves act as a sensor. In [20], they present human sensing in three technical domains: Human sensors online, online social media mining on a large scale and offline human sensors.

- Human sensors online: In this domain, many online marketing applications like Amazon, Flipkart, Paytm are acting as a platform, where people usually interact with these resources. These people provide some query, feedback, opinion as to the form of data. So, we can say that these online users' are the primary source information and acting as a sensor.

- Online social media mining on a large scale: Nowadays, there is a very popular trend of social media like Facebook, Twitter, and LinkedIn, etc. Many users are part of these social media and continuously interchanging their views in the form of opinion, chat, comment, and news. Therefore it is necessary to mine the data of online social media to get some interestingness or surprising information. Fake news detection, opinion mining, crisis response system are the popular result of online social media mining.

- offline human sensors: It is a traditional way of collecting useful information from the people. People usually interact with some organizations (government or non-government) regarding some governmental activities or social camping to share their pieces of informations, feedbacks, queries.

Collective sensor network

Nowadays, Twitter, Flickr, Facebook, Foresquare, etc. are the very famous platform on which individuals, organizations are taking part and do various kinds of activities. The trend of social media and the latest communication technologies introduce a new kind of networking paradigm called a collective network. Sensing through collective network is called collective sensing in which we analyze aggregated informative data coming from aforesaid collective networks. The collective sensor data may be sentiments, comments, user feedback, impressions or perceptions. This platform generates a huge amount of data day by day using several web-based applications like social computing, ubiquitous computing, mobile computing, and collective sensing, etc. Data storage, processing, and abstraction of big data are one of the current key research topics. For example, for handling big data someone can use Lambda architecture [22] and for the context of collecting the sensing platform we can use Map/Reduce framework [23]. These collective sensor data is very useful regarding collect contextual data, accurate measurement through different collective sensor platform, which will provide better analysis and interpret individual to take a better decision.

Some applications for environment sensing

Participatory sensing is better idea for environment sensing because it allows collecting data from various sensing platforms that provide analytic and systematic approaches like understanding patterns, semantics and dynamics of social behavior and many more. Air sense [24], NoiseSense [25] and citizen sense [26] are the applications of environment sensing. Air sense, NoiseSense, and citizen sense are participatory sensing based mobile applications in which data is collected through AQMD (Air Quality Monitoring Device) and smart devices correspondingly. The smart city architecture [12] for a smart city is shown in figure 2.4, using the concept of this architecture we can develop a participatory sensing based framework.

In [24], AQMD is self-sensors, a hardware-based device which is connected through smartphone by WiFi or Bluetooth and data is collected in the cloud. In the cloud, after applying some pre-processing technique and analytical methodologies, the result of the pollution map is generated. Finally, the air pollution map is shown to the end users.

In [25], they have proposed a crowdsourced energy-efficient real-time noise monitoring system called NoiseSense, whose main purpose is to gather context-aware temporal and spatial data from the citizen and generate the individual footprint and generate noise pollution map of the city(using collective data). While in citizen sense [26], it mainly senses the environmental condition using human as a sensor which identifying the road conditions, common health-related issue among citizen and then provide feedback. This idea becomes a very valuable knowledge for both the citizens and the city administrator. The main idea of this application is to discover travel patterns of individual users, crowd flow in urban through their GPS trajectories. This application notifies users about the congestion and current popular places near them.

2.2.2 Participatory Sensing based noise pollution monitoring

In [13], Maisonneuve et. al, present an approach called NoiseTube, whose purpose is to develop a participatory noise pollution monitoring network to measure and map noise pollution with GPS enabled smartphones. The main idea goal of [13] is to turn on GPS-equipped mobile phones into noise sensors in order to measure citizens' personal exposer to noise and their collective noise map in their everyday environment. They use smartphones as an environmental sensor, which senses noise level through integrated sensors like GPS, microphone, motion sensors and many more. They have used the concept of participatory sensing in order to sense the noise of the city. For this purpose, they provide a NoiseTube platform consists of an application. Participants must install this application in their smartphone to turn it into a noise sensor device in order to collect local information from different sensors (noise, GPS coordinates, time, user input) and sends it to the NoiseTube server, where the data is centralized and processed. The end users can visualize their personal exposer to noise and also visualize the collective noise map by aggregating noise measurements collected by the public.

In [14], Rana et. al. propose an end-to-end participatory urban noise mapping system called Ear-Phone. In this paper, they provide a noise map from incomplete and random samples collected by participants. Apart from the noise map, this paper also addresses the challenge of collecting accurate noise pollution reading in a mobile device. For this purpose, they use different mobile phones like Nokia N95 and HP iPAQ mobile devices. In their Ear-Phone architecture, they store the records in form of GPS coordinates, later the communication manager at the central server converts these GPS coordinates of the record to the Military Grid Reference System (MGRS) and store this information in a data repository. When any kind of user query occurs for a particular location, the location information (eg. street, bus stand, airport) of the query is also resolved into MGRS grid indices and the matching grid indices are fetched from the data repository. Before displaying the result of noise map (i.e. noise levels overlaid on a geo-centric Internet map) to the end user, the resultant grid indices have been converted back to GPS coordinates and the corresponding noise levels are overlaid. Later in [15], Rana et. al., propose another method of Ear-Phone whose main purpose is to generate a context-aware noise map using smartphones based on participatory urban sensing. [15] illustrates the use of different interpolation and regularization methods (such as linear interpolation, Gaussian process modeling, and nearest neighbor interpolation and l_1-norm minimization) to reconstruct the urban noise pollution map from incomplete and random noise samples obtained by crowd-sourcing (participatory) data collection. For this purpose, they develop a classifier (Automatic Context Switch Detection [15]) whose main task is to accurately determine the phone sensing context.

Becker et. al. [16] design and developed a smartphone application called WideNoise, which can be used to measure noise by involving citizens in environmental monitoring. The purpose of the application WideNoise is to record both objective (noise samples) and subjective (opinions, feelings, etc.) data. Based on the users' input, an analysis of emerging awareness and learning is performed. This application is feasible for only on Android [18] and iOS [19] platform. This application collects informative data mainly in two parts: the noise sampling component and the perception tagging. For the noise sampling component, users are asked to guess the noise level through a slide bar (decibel scale mapped). The user can also text tags to the noise. In the application's background, the original value of noise levels has been recorded by sensing through a microphone. Once the tags have applied, the mobile application sends informative data to an application server. The application server collects the acquired data and shows the corresponding information (i.e. noise pollution) on a map. In [17], Zheng et. al. infers the fine-grained noise situation noise situation (consisting of a noise pollution indicator and the composition of noises) of different times of day for each region of New York City (NYC), by using the 311 complaint data together with social media, road network data, and Points of Interests (POIs). The data collected using this concept will be a result of "human as a sensor" and "crowd sensing". They implement a three dimensional model for the noise situation of NYC. The three-dimensional components are regions, noise categories, and time slots. They also handle the missing entries of the tensor through a context-aware tensor decomposition approach. They provide a noise pollution indicator for each region in a time span and a noise category. This information can be helpful in making an official decision and aware citizens from noise pollution.

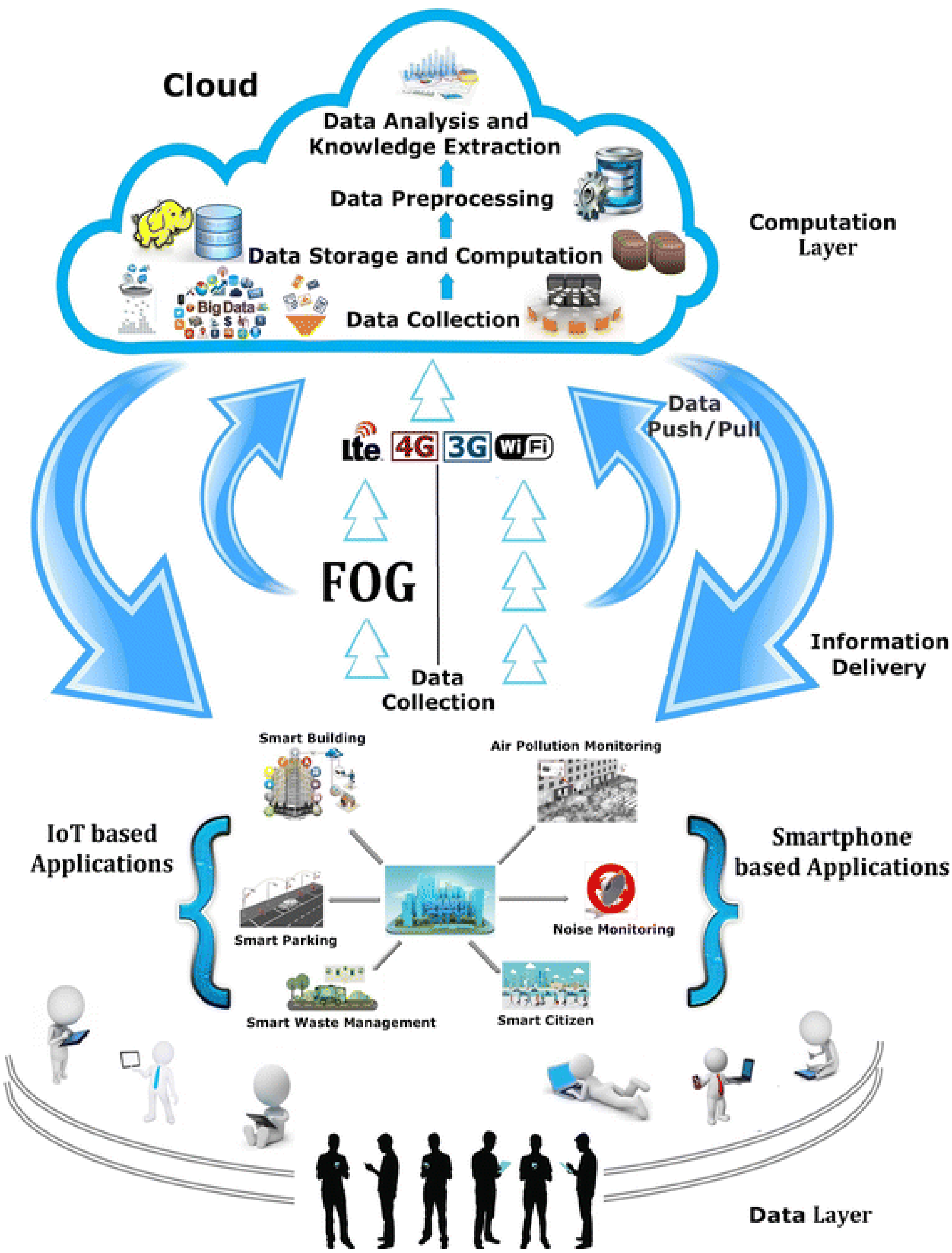

FIGURE 2.4: Smart city architecture. Source: [12]

3 Spatial Interpolation Techniques

3.1 Spatial Interpolation

The word "Spatial" refers to something that occupies some space or in other sense it is related to geometrical space on the earth surface. Interpolation is a technique that predicts the unknown value by using some form of the weighted average of the values at surrounding points or samples. In another way, interpolation can be defined as the process of estimating unknown values that lie between known values. Thus, we can define spatial interpolation as a technique, that predicts/estimate the unknown value of a geographical location by using some form of weighted average method which includes known values of surrounding sample points. The surrounding sample points are also known as control points. Control points are the known points that provide the data essential for the development of interpolator for spatial interpolation. Interpolator may be any kind of mathematical equation. The rationally behind this method is to represent the general trend of any kind of surface. For example, this method is used for estimating the concentration of gold in the earth surface, predicting the spatial variability or predicting/inferring noise value at a particular location. The interpolator is developed by using the data of control points, so the number and distribution of these points may greatly influence the accuracy of spatial interpolation. Interpolation is basically based on the assumption that spatially distributed objects are spatially correlated. The rationale behind spatial interpolation is Tobler's first law of geometry which is states as "Everything is related to everything else, but near things are more related than distant things".

3.1.1 Type of Spatial Interpolation

This method can be classified in many ways. The first way is to classify this method as global and local interpolation. In the case of global interpolation, it basically uses all the control points to estimate the unknown value of a location. While local interpolation uses only those control points which are nearer to the point whose value will be predicted. The main motive of the local interpolation method is to capture the local or short-range variation.

The second way is to classify the spatial interpolation method in the form of exact and inexact interpolation. When the predicted value is the same as the originally known value of that location, it will be considered as exact interpolation. While in the case of inexact interpolation, it predicts approximate value nearer to original value at a particular location. For example, if the observed noise at a particular location in urban is 70 dB, then in the case of exact interpolation, the point representing the particular location on the resulting grid will still have a noise of 70 dB. But, in the case of inexact interpolation, it will predict a different noise value which will not same as the original value but approximately nearer to that observed original value.

The third way is to classify the spatial Interpolation method as deterministic or stochastic. In the case of deterministic interpolation, not any kind of error estimation is required with the predicted value. While stochastic interpolation method provides an estimation of predicting error with estimated variance.

3.1.2 Global Method:

Here we present two types of global methods, which are trend surface models and regression models. In this section, we will only describe the abstract definition of these methods. Trend Surface models are an inexact method. This method generates a polynomial equation using sample points, moreover using polynomial equation it develops a trend surface which is used for estimating the unknown values of points. The mathematical equation or the interpolator which is used to estimate values for other points is represented in equation 3.1.

$$Z_{x,y} = b_0 + b_1 x + b_2 y \tag{3.1}$$

Where the attribute value z is a function of x and y coordinates. The b coefficients are estimated from the known points.

While the regression model is a statistical method that is used to relate dependent variables in a linear equation or an interpolator. This interpolator is then used for predictions or estimation of unknown values.

3.1.3 Local Method

There is a lot of local methods such as Thiessen Polygons, Density Estimation, Inverse Distance Weighted (IDW), Spline and Kriging. But, in this section, we are mainly describing only IDW and Kriging methods.

Inverse Distance Weighted (IDW):

It is an exact and one of the simplest method among spatial interpolation methods. This method is widely used because of its simplicity. It is based on the assumption that sample points that are close to one another are more similar than those that are far away. For the prediction of value for any unmeasured location, this algorithm will use the average of measured values surrounding the prediction location. According to its assumption, measured values nearer to the prediction location will have a greater influence on the predicted value than those farther away. So, the working principle of this method is a local influence and the impact of each measuring point diminishes with distance. It assigns greater weighting value to the point close to the prediction location than those which are far apart and that's why the name is inverse distance weighted. The basic form of predicting unmeasured location value u at a given point x based on measured samples $u_i = u(x) for I = 1,2,\ldots,N$

using IDW is given below:

$$u(x) = \begin{cases} \dfrac{\sum\limits_{i=1}^{N} w_i(x) u_i}{\sum\limits_{i=1}^{N} w_i(x)}, & \text{if } d(x, x_i) \neq 0 \text{ for all i} \\ u_i, & \text{if } d(x, x_i) = 0 \text{ for some i} \end{cases} \tag{3.2}$$

Where

$$w_i(x) = \frac{1}{d(x, x_i)^p} \tag{3.3}$$

w_i is a simple IDW weighting function, which is defined by Shepard [27], x denotes an interpolated(arbitrary) point, x_i is a measured point (interpolating point), $d(x, x_i)$ is distance between measured point x_i to unmeasured point x, N is the total number of measured point used in interpolation and p (positive real number) is the power parameter. The value of p affect the weighting factor. As the value of p increases, the influence of close measured points on interpolated point also increases.

There are lots of different modification on Shepard weighting method, one of them is mention below in equation 3.4, also called as Modified Shepard method.

$$w_i(x) = \left(\frac{max(0, R - d(x, x_k))}{R * d(x, x_i)} \right)^2 \tag{3.4}$$

This weighting method is differ from previous one because it only consider those interpolating point which are lies within R-sphere to the interpolated point.

Kriging

Kriging can be either exact or inexact depending upon its usability. This method differs from simple IDW because it interpolates unmeasured location based on both inverse distances as well as spatial autocorrelation. IDW can be referred to as a deterministic interpolation method because it's interpolator estimate the value based on surrounding measured values which determine the smoothness of the resulting surface. While kriging is a geostatistical method, which is based on a statistical model that includes autocorrelation. Autocorrelation can be defined as the statistical relationships among the measured points. So using the concept of autocorrelation technique, this method is not limited to only produce a prediction surface but also provide some measure of the certainty or accuracy of the predictions.

This method assumes that the distance or direction between measured sample points reflects a spatial correlation, which can be further used for explaining variance in the surface. This method is effective when measured data have some kind of spatial correlation with respect to distance or direction.

The first step of this method is to calculate the spatial correlation of measured sample points, called empirical semivariogram. For this purpose, we will semivariance to measure the correlation of the sample points. A semivariogram is a plot between lag or distance and semivariance among measured points. It basically provides a clear view of the spatial autocorrelation of the measured sample points. The computation of semivarigram for measured sample points is given in equation 3.5 [29].

$$\widehat{\gamma}(h) = \frac{1}{2N(h)} \sum_{1}^{N(h)} (Z(x_i) - Z(x_i + h))^2 \tag{3.5}$$

Where,$\widehat{\gamma}(h)$ is the estimated semivariance at a separation distance h, and $Z(x_i)$ and $Z(x_i + h)$ are known values at two points x_i and $x_i + h$ that are separated spatially by distance h. The term N(h) represents the number of pairs of sample points that are separated spatially by the distance h.

There is several components in a semivariogram which are shown in figure 3.1.

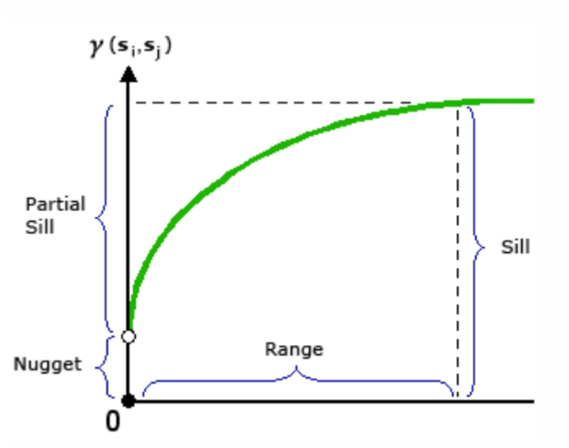

FIGURE 3.1: Illustration of Range, Sill, and Nugget components in a semivariogram. Source: [41]

Lag basically refer as an interval wise distance measure. Range, sill, and nugget are generally used to describe a semivariogram. In figure 3.1, the model is level out at a certain distance. The distance where the model first flattens is referred to as range. The value on the y-axis where the semivariogram reaches the range is termed as the sill. Sometimes the model depicts a partial sill which the sill minus the nugget. Theoretically, the value of the semivariogram should be zero at zero separation. However, the value of the semivariogram is greater than zero when at an infinitely small separation distance. It means that the semivariogram often exhibits a nugget effect which is nothing but intercept on the y-axis. This effect can be used to measure errors or spatial sources of variation at distance smaller than the sampling interval (or both).

After computing the empirical semivariance of measured sample points, it then fits the appropriate model which satisfies the spatial autocorrelation or semivariance of measured sample points. The choice of model depends on prior knowledge of the phenomena and the nature of semivariance. There is a lot of correlation models such as a linear semivariance model, circular semivariance model, spherical semivariance model, exponential semivariance model, and Gaussian semivariance model which is given below:

- **linear semivariance model:** The mathematical equation of linear semivariance model is given in equation 3.6. As the name suggests, the nature of semivariance is linear. It is applied when spatial autocorrelation of measured sample points is decreasing linearly with increasing distance. But after some distance the growth of this curve becomes constant. This model is shown in figure 3.2.

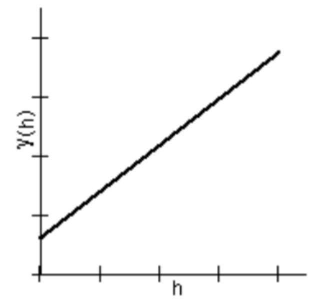

FIGURE 3.2: A graphical representation of Linear semivariance model. Source: [41]

$$\gamma(h) = \begin{cases} C_o + C\left(\frac{h}{a}\right), & \text{for } 0 < h \leq a \\ C_o + C, & \text{for } h > a \\ 0, & \text{h = 0} \end{cases} \tag{3.6}$$

where C_o is the nugget variance, C is the a priori variance of the autocorrelated structure and a is the range of the structure.

- **Circular semivariance model:** The general shapes of circular semivariance model is shown in figure 3.3. This model is applicable where the progressive decrease of samivariogram circular in nature until some distance, beyond which autocorrelation is zero. Equation 3.7 describe the mathematical form of this model.

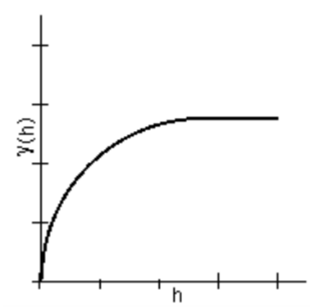

FIGURE 3.3: A graphical representation of Circular semivariance model. Source: [41]

$$\gamma(h) = \begin{cases} C_o + C\left(1 - \frac{2}{\Pi} cos^{-1}\left(\frac{h}{a}\right) + \sqrt{1 - \frac{h^2}{a^2}}\right), & \text{for } 0 < h \leq a \\ C_o + C, & \text{for } h > a \\ 0, & \text{h = 0} \end{cases} \tag{3.7}$$

where C_o is the nugget variance, C is the a priori variance of the autocorrelated structure and a is the range of the structure.

- **spherical semivariance model:** It is one of the commonly used model. It is applicable when autocorrelation of measured sample points decreases progressively. After some distance the autocorrelation become zero. The nature of this model is shown in figure 3.4 and mathematical equation is given in equation 3.8.

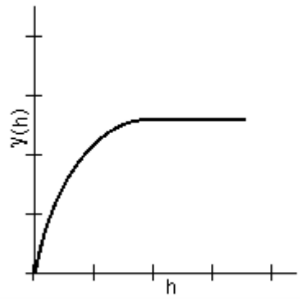

FIGURE 3.4: A graphical representation of Spherical semivariance model. Source: [41]

$$\gamma(h) = \begin{cases} C_o + C(\frac{3h}{2a} - \frac{1}{2}\frac{h}{a}^3), & \text{for } 0 < h \leq a \\ C_o + C, & \text{for } h > a \\ 0, & \text{h = 0} \end{cases} \tag{3.8}$$

where C_o is the nugget variance, C is the a priori variance of the autocorrelated structure and a is the range of the structure.

- **Exponential semivariance model:** This model is applicable only if the autocorrelation is decrease exponentially with increasing distance. Autocorrelation disappears completely only at an infinite distance. The mathematical equation of this model is given in equation 3.9 and pictorial representation is shown in figure 3.5.

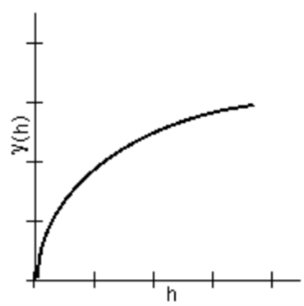

FIGURE 3.5: A graphical representation of Exponential semivariance model. Source: [41]

$$\gamma(h) = \begin{cases} C_o + C\left(1 - exp\left(\frac{-h}{a}\right)\right), & \text{for } h > a \\ 0, & \text{h = 0} \end{cases} \tag{3.9}$$

where C_o is the nugget variance, C is the a priori variance of the autocorrelated structure and a is the range of the structure.

- **Gaussian semivariance model:** The mathematical equation is given in equation 3.10 and the auto correlational behaviour of this model is shown in figure 3.6.

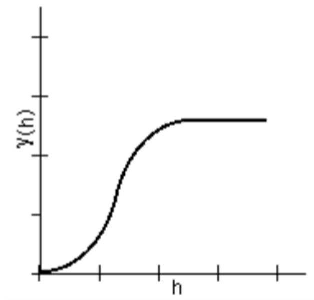

FIGURE 3.6: A graphical representation of Gaussian semivariance model. Source: [41]

$$\gamma(h) = \begin{cases} C_o + C\left(1 - exp\left(\frac{-h^2}{a^2}\right)\right), & \text{for } h > a \\ 0, & \text{h = 0} \end{cases} \tag{3.10}$$

where C_o is the nugget variance, C is the a priori variance of the autocorrelated structure and a is the range of the structure.

These models are used to predict the unmeasured location. Like IDW interpolation, Kriging also forms weights from nearby measured values of the interpolation point to predict values. In IDW interpolation, the measured values closest to interpolation point have greater influence. However, the weight of kriging is more sophisticated than those of IDW. In the case of IDW, weights are computed from a simple algorithm based on distance, but in the case of kriging, weights are determined with the help of a semivariogram. The mathematical equation of kriging is given in equation 3.11 [30].

$$\widehat{z}(x_0) - \mu_0 = \sum_{1}^{k}(\lambda_i[z(x_i) - \mu(x_0)]) \tag{3.11}$$

Where μ_0 is called stationary mean which is computed as the average of the whole data set [28], λ_i is kriging weight; k and $\mu(x_0)$ are the total number of observation points and the mean of the observation points within the search neighbourhood.

There is a different form of kriging that exists in its literature [31]. The three mostly used kriging methods are simple kriging, ordinary kriging, and universal kriging.

1. **Simple kriging:** As the name suggests simple kriging is mathematically very simple to represent. But this variant of kriging is the least general. It assumes that the mean of measured sample data is known i.e, in another way the expectation of the random field to be known, and relies on a covariance function. In this variant, the equation 3.11 is modified by replacing $\mu(x_0)$ with stationary

mean μ_0. The equation for simple kriging is given in equation 3.12.

$$\widehat{z}(x_0) - \mu_0 = \sum_1^k (\lambda_i [z(x_i) - \mu_0]) \tag{3.12}$$

2. **Ordinary kriging:** Ordinary kriging is another variant of kriging. In this case, it assume that there is not any kind of drift in the spatial variation in known measured points. The equation 3.11 is modified by replacing μ_0 by $\mu(x_0)$, which is the local mean of observing measured sample data within search neighbour. By using this concept it basically capture the local effect of sample data and the predicting value get influence by local sample measured data near the interpolating point. It focuses on the spatially correlated component and uses the fitted model of semivariogram directly for interpolation. The modified form of kriging i.e, ordinary kriging is given in euation 3.13.

$$\widehat{z}(x_0) - \mu(x_0) = \sum_1^k (\lambda_i [z(x_i) - \mu(x_0)]) \tag{3.13}$$

3. **Universal kriging:** Universal kriging (UK) is another variant of kriging in other word is is an extension of Ordinary kriging which assumes that the spatial variation in known measured points has a drift or trend in addition to spatial correlation between sample points.

4

Participatory sensing based framework

4.1 Overview

We have designed a participatory sensing based framework inspired by smart city architecture [12] which senses the noise and generate noise pollution map of the urban. The pictorial representation of our framework is shown in figure 4.1. Our designed framework consists of mainly three parts, which are noise sensing, pre-processing and service. At first, the measurement of noise is done via smart devices like smartphones and tablets and then transmitted to the cloud server. These measured noise samples are collected from different locations at different timestamps in the urban. So it becomes an important criterion to combine these noise samples collected by different users. Cloud server combines spatiotemporal based measured noise samples which are collected by different participants (individuals, volunteers, and participant or citizen).

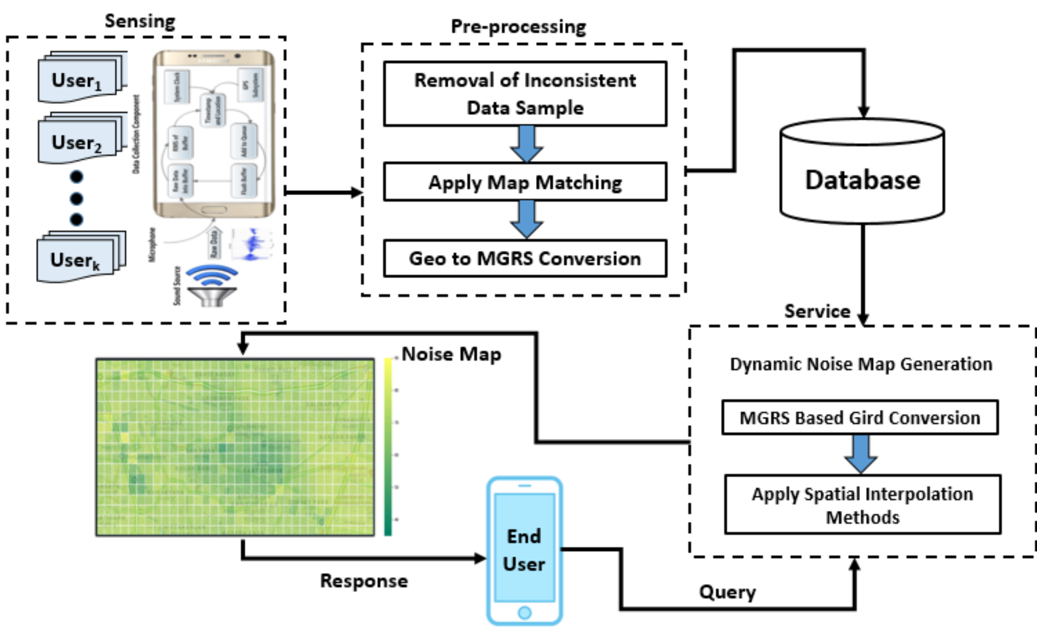

FIGURE 4.1: Our participatory sensing based designed framework

Sensing and data collection

Sensing and data collection is one of the important part of participatory sensing based framework. We use NoiseSence [25] application of this purpose. NoiseSense is an android based application suitable for those smart devices having an android platform. Users must install this application in their smart devices. The snapshot of the internal processing of this application is shown in the sensing part of the figure 4.1. Before sensing the noise data users need to calibrate their smart devices (see section 5.2). We can sense both objective and subjective data which is sensed by users. Objective data consists of noise measurement, GPS location, Timestamp, 3-axis accelerometer, and proximity sensor values. We can also extract information about the footprint and individual noise of the user using their GPS coordinates and timestamp values. While in the case of subjective data sense, it consists of contextual information of the location which is inputted by users. Using the NoiseSense application, data gets recorded at an interval of every 5 seconds in the format < *user id, smartphone model, latitude, longitude, timestamp, noise level,* acc_x, acc_y, acc_z, *proximity* > and are uploaded to a database. These context-aware noise data is very useful for data cleaning and pre-processing processes.

Pre-processing

The pre-processing steps our designed framework consists of multiple stages. The first step of this phase is to handle with inconsistent data. For this purpose we simply ignore the missing sample values and also check whether noise sample is collected in ideal condition or not like: using 3-axis accelerometer we can check whether the smart devices were inside (pockets or bags) or outside (hands or nearby users), using the value of proximity sensing [15] we can check whether devices are ideal or not. So observing the value of a 3-axis accelerometer and proximity sensor we can clean the inaccurate data samples. Using the value of sensors like accelerometer, gyroscope, light sensor and the proximity sensor (use as features), we can also classify an accurate noise measurement by applying a machine learning algorithm like Naive Bayesian [35] or decision tree [35] etc. The second phase is that we apply a map matching algorithm [32] to correctly fit the noise sample data on GPS coordinates. In the third phase, we convert the GPS coordinates (latitude and longitude) into corresponding MGRS (Military Grid Reference System) coordinates or grid index. Using the MGRS grid technique we can divide the region into smaller grids.

Service

After the pre-processing, we can apply our proposed spatial interpolation methods in order to generate the noise pollution map of the city. We have proposed a variety of methods and each method has its own signification. The end user may query regarding the noise pollution map. Our approach is to provide them a suitable noise pollution map. The noise pollution map may vary user to user. For example, some user is only interested in their personal footprint pollution map, some of them may be interested in pollution for the particular location (Region of Interest), or some of them may ask for entire pollution map of the region. So based on the user's query we will provide a noise pollution map or we can also aware citizens from noise pollution and recommend some other kind of services (e.g., recommending suitable path) which will be helpful in improving the lifestyle of urban.

4.2 Noise level measurement

Sound pressure level (SPL) can be used to measure the noise level. SPL is the logarithmic RMS value of the sound or acoustic pressure caused by sound wave relative to a reference value in decibels (dB). The formula that is used to calculate the SPL are given in below:

$$SPL(dB) = 10log_{10}\left(\frac{p_A}{p_O}\right)^2 \tag{4.1}$$

Where, p_A is the acquired sound pressure in Pascal (Pa) unit area in meters and p_O is reference sound pressure level (typically 20 μPa) in Pa. But, the amplitude of sound pressure level (SPL) is varies over time interval. So it is better to measure an equivalent technique that measures the continuous SPL. For this purpose we used to measure an equivalent continuous sound level (Le) whose result is expressed in units of decibels. Le is used to represent an imaginary constant SPL that is used to measure SPL in a given interval. In other word we would say that it results the same energy as the fluctuating sound level we are measuring over a given time interval. The mathematical equation for calculation of L_{eq} is given below:

$$L_{eq} = 10log_{10}\left[\frac{1}{t_{end} - t_{start}}\int_{t_{start}}^{t_{end}}\left(\frac{p_A}{p_O}\right)^2 dt\right] \tag{4.2}$$

Where, t_{start} and t_{end} are start and end of the time interval respectively, p_A and p_O are discussed in equation 4.1. But, the noise level measured by the equation 4.2 may not gives correct results. So, we need to calibrate smartphones which is discussed in section 5.2.

4.3 GPS to MGRS Conversion and grid conversion

To plot the noise pollution map of the region or entire city some kind of visualization is required. We divide the entire region into small reference grids (equal size) for the analysis of participatory sensing noise data. Dividing the region into small reference grids is helpful for visualization of data sample like from which region (in our case grid) data sample is collected. It also helps in inferring the noise for unmeasured locations because it seems impractical to infer noise at every GPS location. Instead of the GPS coordinate system, we should use some kind of a grid reference technique. So in our work, we use the MGRS [33] technique which simply converts the GPS coordinates into corresponding MGRS grid index based on their precision level. In our work, we use 1 meter of precision i.e. each noise data sample is represented by a 1×1 square meter grid. We can also use some other reference systems like Universal Transverse Mercator (UTM), Universal Polar Stereographic (UPS) [34], etc., but these reference systems are quite complex and some modification might be required. An MGRS co-ordinate or grid reference consists of three parts: grid zone designator, the 100,000-meter square identifier and a numeric location (easting and northing). Figure 4.2 shows the MGRS coordinates of the part of our study region (Jadavpur University, Kolkata). For dividing the region into a smaller reference grid, we use a 100×100 square meter grid. After applying our spatial interpolation methods, each grid is associated with a noise value which may be either measured noise or inferred noise.

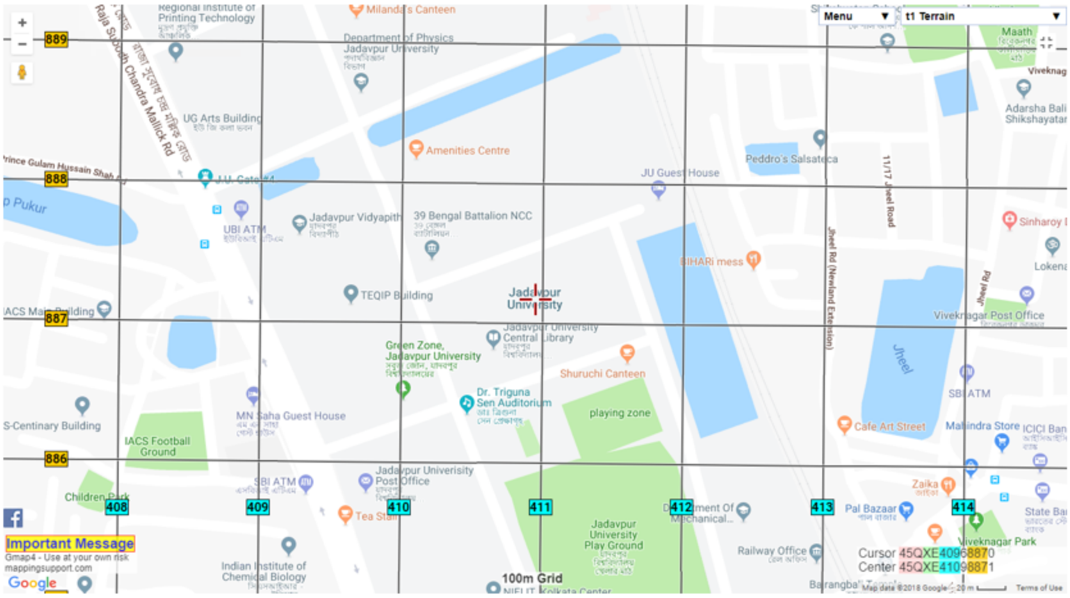

FIGURE 4.2: MGRS coordinates of Jadavpur University, Kolkata, India.

4.4 Grid-MSM

In our work, the first proposed method is an MGRS based local spatial interpolation technique, called Grid-based Modified Shepard's Method (Grid-MSM). Grid-MSM predicts the noise of grids locally using the weighting technique of Modified Shepard's Method. Modified Shepard's Method is nothing but the modified version of the inverse distance weighted (IDW). As we already discussed in section 4.3 that predicting noise value for every GPS location seems impractical, so we develop a grid-based method for noise prediction. Thus, the Grid-based method is one of the suitable methods for inferring noise value.

The procedure of Grid-MSM is mention in algorithm 4.4. Algorithm 4.4 takes a set of inputs namely sample noise data, radius, number of neighbour and dimensions of each grid. The first step of Grid-MSM is to convert geo coordinates (latitude, longitude) into corresponding MGRS grid reference and then divide the region into a set of smaller grids having dimension $D \times D$, where D is the dimension of each grid. In the next step, it simply constructing a set NN_i of sample points that are within a radius $\Re$ from the midpoint of the considered grid. After that, it checks whether the absolute value of set NN_i is meeting some threshold NB_{Th} or not. If $|NN_i| \geq NB_{Th}$ then it consider only NB_{th} number of nearest neighbour for noise prediction of that grid. The optimum value of NB_{Th} and dimension D is shown in result section. After that, it calculate the weights of closet nearest neighbours (NB_{Th}) using equation 4.3.

$$w_{ij} = \begin{cases} \left[\frac{max(0,\ \Re - d(mid,\ s_{ij}))}{\Re * d(mid,\ s_{ij})} \right]^2, & \text{if } d(mid, s_{ij}) \geq 1 \\ 1, & \text{if } d(mid, s_{ij}) < 1 \end{cases} \tag{4.3}$$

Here, w_{ij} is the weight of j_{th} neighbour of i_{th} grid, $\Re$ is the specified radius, mid is the midpoint coordinates of grid g_i, s_{ij} is the j_{th} neighbour of i_{th} grid, and $d(mid,\ s_{ij})$ denotes the euclidean distance between the midpoint of i^{th} grid and midpoint of its

j^{th} neighbour.
Next, it predicts the noise value of g_i based on the Modified Shepard's Method which is given in the formula 4.4.

$$GN_{pred_i} = \begin{cases} \dfrac{\sum\limits_{s_{ij} \in NN_i} w_{ij} \times noise[s_{ij}]}{\sum\limits_{s_{ij} \in NN_i} w_{ij}}, & \text{if } NN_i \geq NB_{Th} \\ NULL, & \text{otherwise} \end{cases} \tag{4.4}$$

Here, GN_{pred_i} is the predicted noise value of the i_{th} grid, NB_{Th} is the predefine threshold value for nearest neighbour, NN_i is the set of nearest neighbour of the grid g_i, $noise[s_{ij}]$ represents noise value of j_{th} neighbour of the i_{th} grid.
But, if the checking condition $|NN_i| \geq NB_{Th}$ is false, then it simply puts $NULL$ to the noise value for that grid. The $NULL$ indicates that there is an insufficient number of the nearest neighbour of that grid. Finally, this algorithm returns a set of grids G along with their predicted noise value GN_{pred}.

The procedure of Grid-MSM algorithm 4.4:

- **Step 1.** Convert longitude and latitude coordinates into MGRS coordinate for each Sample data.
- **Step 2.** Create a set of grids G having dimension $D \times D$.
- **Step 3.** Repeat the step 4 to step 8 for each grid g_i.
- **Step 4.** Compute all nearest neighbours within radius $\Re$ and store in NN_i.
- **Step 5.** Check whether if $|NN_i|$ is greater than a predefined neighbouring threshold NB_{Th} or not. If so, GOTO step 6 otherwise GOTO step 8.
- **Step 6.** Select only NB_{Th} closet nearest neighbours and compute weights for each neighbour using equation 4.3.
- **Step 7.** Predict the noise value for the grid g_i using formula 4.4.
- **Step 8.** Put a NULL value to the noise value for that grid g_i and store this noise value in GN_{pred}.
- **Step 9.** Return a set of grids G along with grids' noise GN_{pred}.

This method is very simple and easy to implement. however, it suffers from some limitations. The most common limitation is that it is unable to predict the noise value of grids when data samples are sparse (i.e. unavailability of the data samples in the neighborhood of the grid). This algorithm does not reuse the predicted noise value as a data sample for the prediction of remaining girds. To overcome aforesaid limitations we have proposed another method I-Grid-MSM, and Grid-MSM-SK. These methods have the ability to generate a better noise map. However, these methods have their own merits and demerits depending upon several situations (see section 5.4.1).

4.5 I-Grid-MSM

I-Grid-MSM (Iterative Grid-MSM) is an iterative version of Grid-MSM. This is another type of interpolation method in which it reuses the predicted noise values when data samples are sparse. This method has the ability to propagate the local effect of grids noise to their neighbours in an iterative manner. Due to iterative propagation of the grid's noise, I-Grid-MSM has the ability to predict each and every grid's noise.

The procedure of I-Grid-MSM is described in algorithm 4.5. The inputs of this algorithm are sample data, radius for Grid-MSM, number of neighbours for Grid-MSM, A chess board distance CBD for iterative computation, Threshold $Grid_{Th}$ and dimension of each grid. The first step of this algorithm is to compute the local grids noise value by using the Grid-MSM method. As mention is section 4.4, Grid-MSM is able to predict only those grid having a sufficient number of sample data. Now our task is to predict the rest of the grids noise values. In the first iteration, for each grid g_i, we check that whether the value of noise grid g_i is predicted or not. If not then we calculate the nearest neighbour using chessboard distance CBD. Next, this algorithm construct a set GNN_i of nearest neighbour $\in GNN_{total}$ having noise value not $NULL$. Further, compute the value of Θ and check whether it meets the threshold condition or not. If so, we compute weight w_{ij} for each grid neighbour $g_i \in GNN_i$ using equation 4.5.

$$w_{ij} = \left[\frac{max(0,\ (\sqrt{2}CBD) - d(mid_i,\ mid_j))}{(\sqrt{2}CBD) * d(mid_i,\ mid_j)}\right]^2 \tag{4.5}$$

Here, w_{ij} represents the weight value of j_{th} grid of i_{th} neighbour, CBD is a chessboard distance, $d(mid_i,\ mid_j)$ is the distance between midpoint of i^{th} grid and midpoint of j^{th} grid.

Next, this method computes the predicted noise value GN_{pred_i} of grid g_i using the given formula 4.6 which is based on the weights w_{ij} and corresponding noise value GN_{pred_j}.

$$GN_{pred_i} = \frac{\sum\limits_{g_j \in GNN_i} w_{ij} \times GN_{pred_j}]}{\sum\limits_{g_j \in GNN_i} w_{ij}} \tag{4.6}$$

Here, GNN_i is a set of all nearest grid neighbour of g_i whose predicted noise values would not be NULL.

But if Θ not meets threshold condition $Grid_{Th}$, then noise value remains same (i.e. $NULL$ value). If the noise value of any grid is not computed then we go to the next iteration. The outputs of the previous iteration become the inputs of the next iteration. The above-mentioned process continues until noise values for all grids have been computed.

The procedure of Iterative-Grid-MSM algorithm 4.5:

- **Step 1.** Compute only those girds' value having sufficient number of sample points using Grid-MSM algorithm.
- **Step 2.** Stop and GOTO step 8 when noise values of all the grids have been computed otherwise GOTO step 3.
- **Step 3.** Repeat the step 4 to step 7 for each grid g_i.
- **Step 4.** If $GN_{pred_i} == NULL$ then GOTO next step otherwise GOTO step 8.
- **Step 5.** Create a set of grid neighbours GNN_{total} having a chess board distance CBD from grid g_i and also create a set of all nearest grid neighbour GNN_i of g_i having predicted noise value not NULL. Compute $\Theta = \frac{|GNN_i|}{|GNN_{total}|}$.
- **Step 6.** If Θ is greater than a predefined threshold $Grid_{Th}$ then we compute weights values for all grids neighbours GNN_i using equation 4.5 otherwise GOTO step 3.
- **Step 7.** Predict the noise value for the grid g_i using formula 4.6 and store this result in GN_{pred_i}.
- **Step 8.** GOTO step 2 if all grids have been traversed for i_{th} iteration otherwise GOTO step 3.
- **Step 9.** Return a set of grids G along with grids' noise GN_{pred}.

Chess board distance: It is the maximum distances that we cover along x direction and y direction. D_8 distance or chess board distance is defined as

$$D_8(p,q) = max(|x - s|, |y - t|) \tag{4.7}$$

s={ $q \mid D_8\ (p,\ q)\ \leq\ r$} forms a square centred at p.
Points with $D_8 = 1$ are 8 neighbours of p.
All the point with chess board distance of 1 from a point p they are nothing but 8 neighbours of point p. similarly the set of points with a chess board distance equal to two will be just the points outside the points having chess board distance one.

4.6 Grid-MSM-SK

Grid-MSM-SK refers as Grid-based Local Modified Shepard's Method following Simple Kriging. This method is a kind hybrid spatial interpolation method. Grid-MSM-SK is based on both Grid-MSM and Simple Kriging methodologies. The main concept of this method is to analyze both local and global views of the region. The ideology behind the local view is originated from Grid-MSM which means that we are inferring the grid's noise based on the nearest sample data (local to the particular grid i.e. sample data within a predefined radius $\Re$) from the midpoint of that grid. On the other hand, the ideology behind the global view is coming from grids concept. If the local view fails to predict the noise value of a particular grid then we can use a set of grids (already predicted) to predict the noise value of that grid. So, for the local view analysis, we apply Grid-MSM to predict those grids having a sufficient number of sample data points. The output of Grid-MSM is then used as input for the simple kriging (SK) method. The simple kriging method is used to predict the rest of the grid which is not predicted yet.

The procedure of Grid-MSM-SK describes in algorithm 4.6. The input for the Grid-MSM-SK are sample data, a radius for Grid-MSM, number of neighbours NB_{Th} for Grid-MSM, K is number of neighbour points to consider for simple kring, dimension of each grid, A set of models, Simple kriging bandwidth bw, Simple kriging distance hs. At first, Grid-MSM has been called to compute the local grids noise values. The output of Grid-MSM has been used as input for the simple kriging. The experimental semivariogram has generated based on the output grids values using kriging bandwidth bw and kriging distance hs. The function representation of semivariogram is discussed in equation 3.5 in chapter 4. The next step of this algorithm is to choose the best fit model for the experimental semivariogram. Next, a covariance function is generated using the dataset, $model$, hs, bw. Next, this method computes the global mean of the output of Grid-MSM. For each grid g_i having $NULL$ value, we predict noise value of that grid using a simple kriging method. In the end, the algorithm returns a set of grids G along with their predicted noise value GN_{pred}.

The procedure of Grid-MSM-SK algorithm 4.6:

- **Step 1.** Compute only those girds' value having sufficient number of sample points using Grid-MSM algorithm.
- **Step 2.** Generate a Semivariogram of the grid data set using equation 3.5.
- **Step 3.** Select a model which fits best for experimental semivariogram.
- **Step 4.** Create a covariance function using dataset, model, simple kriging distance hs, and simple kriging bandwidth bw.
- **Step 5.** Calculate mean m of the GN_{pred} having not $NULL$ noise value.
- **Step 6.** Repeat the step 7 to step 10 for each grid g_i.
- **Step 7.** If $GN_{pred_i} == NULL$ then GOTO next step otherwise GOTO step 11.
- **Step 8.** $\mathbf{P} \leftarrow$ select top K' neighbours which is closest to g_i.
- **Step 9.** Calculate covariance matrix $\mathbf{w_1}$ using $model$ and also calculate covariances $\mathbf{w_2}$ between the data points and an unsampled point g_i. Calculate kriging weights using the given formula:
  $\boldsymbol{\lambda} \leftarrow \mathbf{w_1^{-1}} \times \mathbf{w_2}$
- **Step 10.** Predict the noise value for the grid g_i using the given formula:
  $GN_{pred_i} = \boldsymbol{\lambda}\mathbf{P}[noise - m] + m$
- **Step 11.** GOTO step 12 if all grids have been traversed otherwise GOTO step 6.
- **Step 12.** Return a set of grids G along with grids' noise GN_{pred}.

5 Experiments and Results

5.1 Study Area and Data collection strategies

We have collected noise samples near about Jadavpur University, Santoshpur area, and Garfa, Kolkata, India. Our study area covers around 7 square KM (Kilometer). These regions are densely populated having diverse land use. Our study area is shown in figure 5.1.

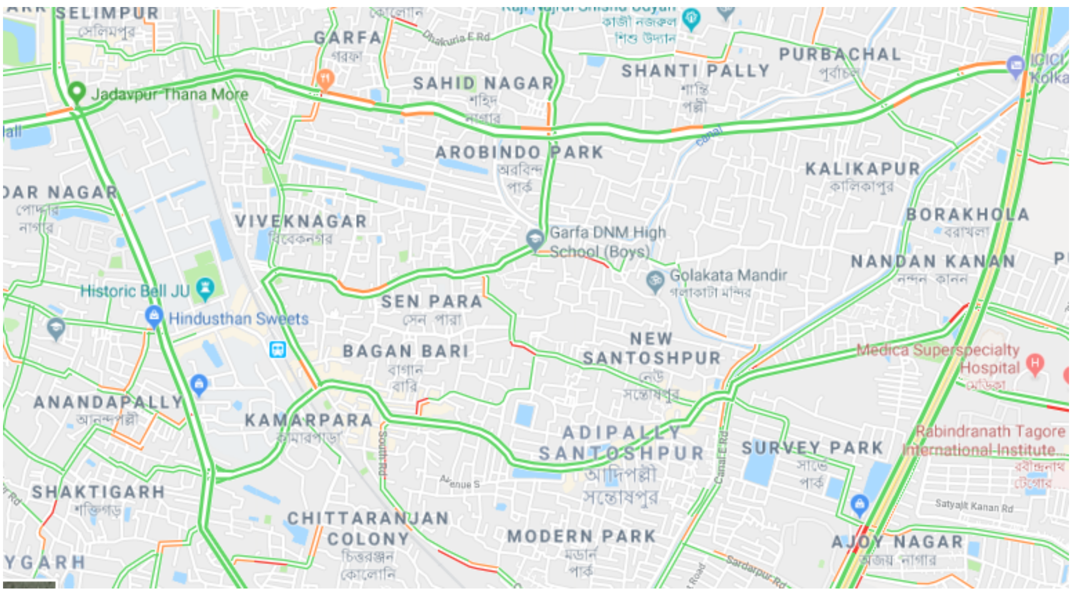

FIGURE 5.1: Google map [38] of our study area. Our study area is a part of south Kolkata near Jadavpur, India.

We have chosen 30 participants (including volunteers, citizen) for the purpose of data collection. Every participant was required to install our application NoiseSense [25] in their respective smartphones (must support Android). In NoiseSense application, data gets recorded at an interval of 5 seconds in the SQLite database in the format $<$ *user id, smartphone model, latitude, longitude, timestamp, noise level,* acc_x, acc_y, acc_z, *proximity* $>$. Apart from this participants can also input some subjective data like Noise conditions, Point of interest and surrounding environment behaviour. Activities of the application NoiseSense for logging ambient noise data are shown in Figure 5. Participants have collected noise data at different locations, different intervals of time throughout the year 2018. We have collected ground truth noise data by using standard noise measuring devices like: Mextech SL-4012 sound level meter [36] and Amprobe SM10 [37].

5.2 Calibrating Smart devices

While dealing with a participatory sensing based framework for data collection strategy, calibration is essential and necessary conditions for smartphones. So for calibration purposes, we need to estimate the calibration factor in order to measure data accurately. To calculate the calibration factor, we measure and compare the readings from SLM (Sound Level Meter) and different smart devices under different sound levels (sound intensity). In our case, we have measured sound level using both devices (SLM and the particular smart devices) in real the environment. We have collected data from a different context like railway station, bus stand, hospital, traffic, streets, etc. The plot between SLM and smart device (Panasonic Eluga) is shown in figure 5.2. Moreover, we use linear regression (least squares fitting technique) to show that there exists a linear relationship between smart device noise level measurements and true noise levels (i.e. SLM readings). It may happen that different smart devices are fitted with different linear models. The general equation for fitting smart devices in linear model is shown in equation 5.1, where SM_x is the smart device reading, α and β are regression coefficients. The value of α and β for Panasonic Eluga are 0.9146 and 48.014 respectively and corresponding R^2 measure is 0.8936. Further, we need to add the difference between linear models of SLM and smart devices in the form of the calibration factor.

$$F_{SM_i} = \alpha SM_x + \beta \tag{5.1}$$

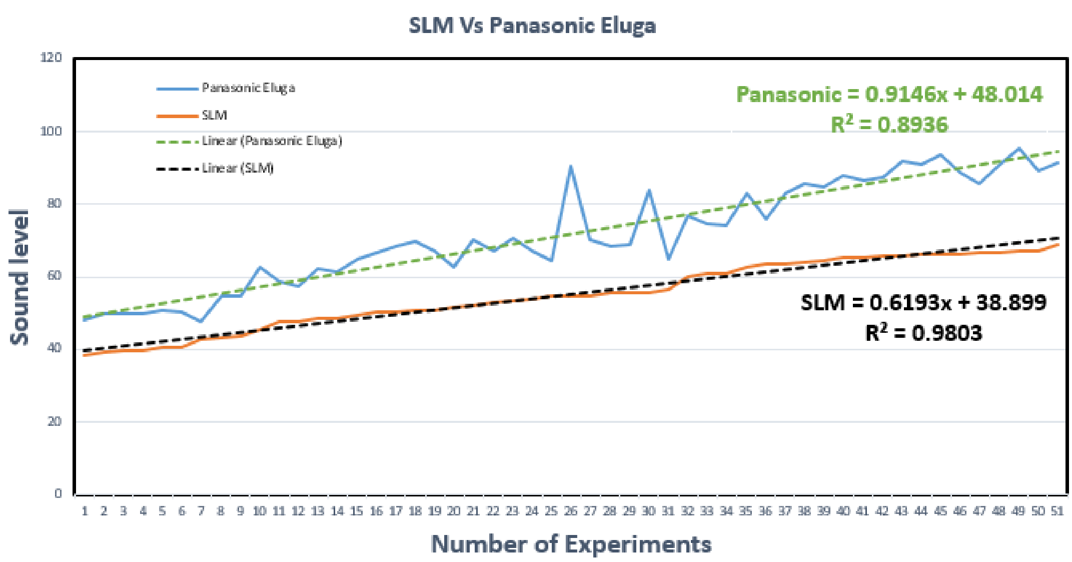

FIGURE 5.2: Plot between SLM (Sound Level Meter) and Panasonic Eluga smartphone.

5.3 Parameters and model selection

The dimension of each grids D and the threshold NB_{Th} for the nearest neighbour for the grid should be predefined before the implementation of Grid-MSM. To find the optimal combination of these two parameters we measure RMSE (Root Mean Square Error) by varying different combinations of NB_{Th} and D. The plot of this combination and respective RMSE measure is shown in figure 5.3. It can be observed that the range of dimensions between 50-80 meters and NB_{Th} is around 10 have an optimal solution, but choosing this combination will result in many grids unpredicted. Moreover, when we increase the number of neighbours corresponding grid dimension range 50-80, it becomes unable to predict noise value of grids due to the shortage of nearest sample points and RMSE value also increases. however, if the range of D is between 80-120 meters and NB_{Th} is around 10, then we observe an appropriate result. Thus, in our work, we choose the dimension of each grid 100×100 and the threshold $NB_{Th} = 10$.

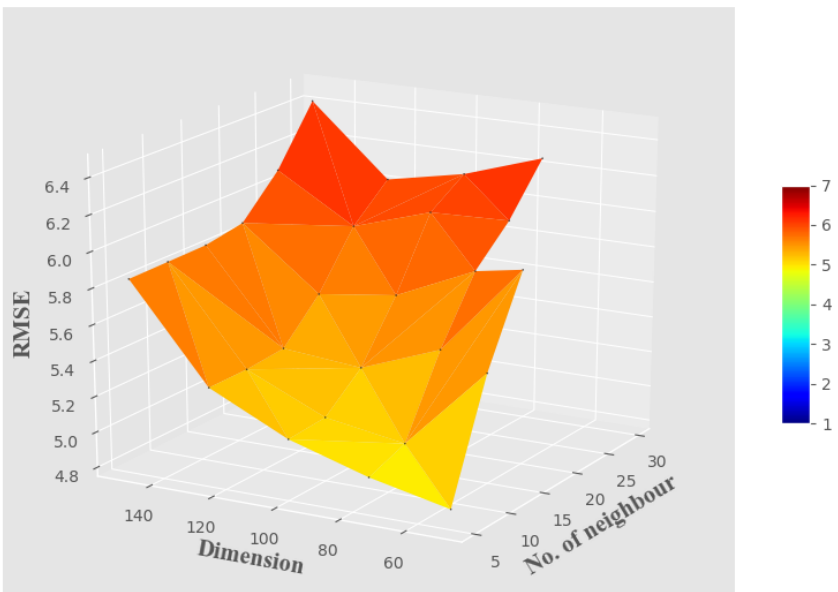

FIGURE 5.3: 3-dimensional plot of prediction accuracy (RMSE values) for different combinations of neighboring threshold NB_{Th} and grid dimension D.

For I-Grid-MSM, we choose chessboard distance $CBD = 1$ for computing nearest neighbours of the unpredicted grid. It is because when we choose $CBD = 2$ it will select surrounding grids neighbours approx 300 meters. It is better to predict long distance iteratively as compared to predict them directly.

The fitting models for Grid-MSM-SK are shown in figure 5.4. We choose a spherical model, which is the best fit as compared to other models for this approach. The parameters of spherical model are: nugget variance $(C_0) = 12.354196$, sill $(C) = 77.421956$ and range $(a) = 8.632653$. We use $K' = 12$ for estimating each grid's noise.

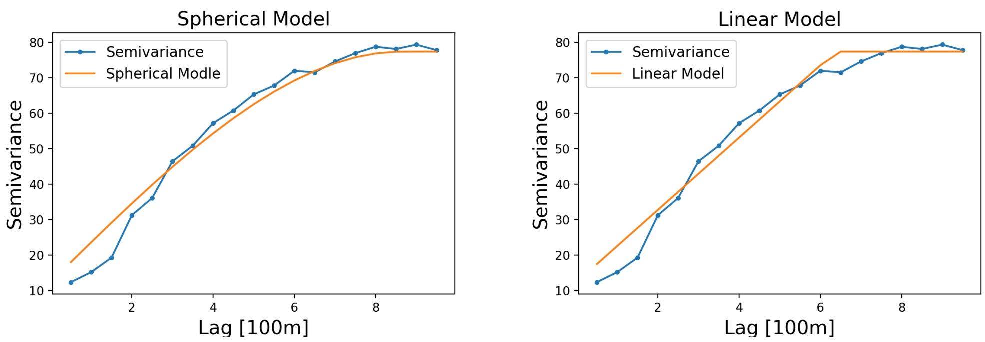

FIGURE 5.4: Model Fitting for the experimental semivariogram.

5.4 Results

5.4.1 Noise Pollution map

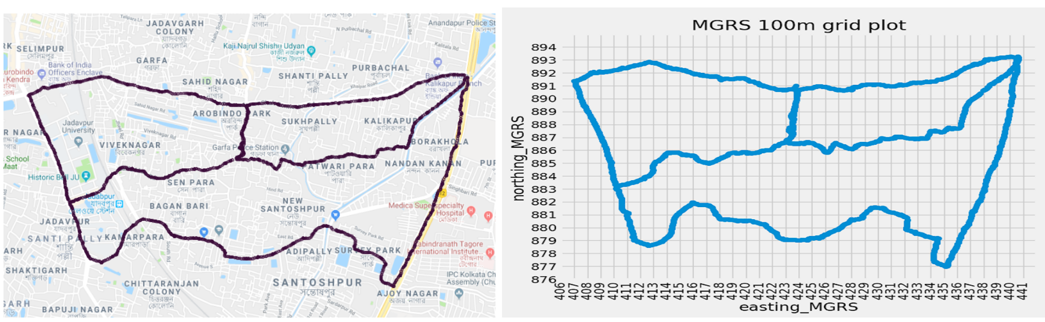

FIGURE 5.5: (a) Noise trajectory plot on google map (b) Grid conversion using MGRS approach.

Noise plot on google map and corresponding MGRS grid map of our study area are shown in figure 5.5 (a) and (b). We have also shown the noise pollution map produced by our proposed methodologies (Grid-MSM, I-Grid-MSM, and Grid-MSM-SK). Figure 5.6 shows that noise pollution map produced by Grid-MSM. The limitation of Grid-MSM is already discussed in section 4.4. Grid-MSM is unable to predict noise value for all grids. However, Iterative-Grid-MSM is able to predict all grids' noise by propagating the local noise to their neighbours. The noise pollution map produced by I-Grid-MSM is shown in figure 5.7. One thing is noted that in the bottom right portion of the figure 5.6 is noisy for predicted girds. By applying I-Grid-MSM iteratively, it simply propagates these noise values to their neighbour's grids. But, those unpredicted grids may not so noisy. Another limitation associated with I-Grid-MSM is about a number of iterations i.e. when the data samples are very sparse and cover a large region, it may take more iterations to predict noise value for all girds. The prediction process of this method does not consider the spatial correlation and other contextual features. Due to this reason, it faces the aforesaid limitations. Later in section 5.4.3, we discuss some key features of this method and compare it with some baseline spatial interpolation methods. Figure 5.8 shows that the noise pollution map produced by Grid-MSM-SK. These methods have used autocorrelation for predictions. The map produced by Grid-MSM-SK predicts grids' noise near about global mean (mean of already predicted girds).

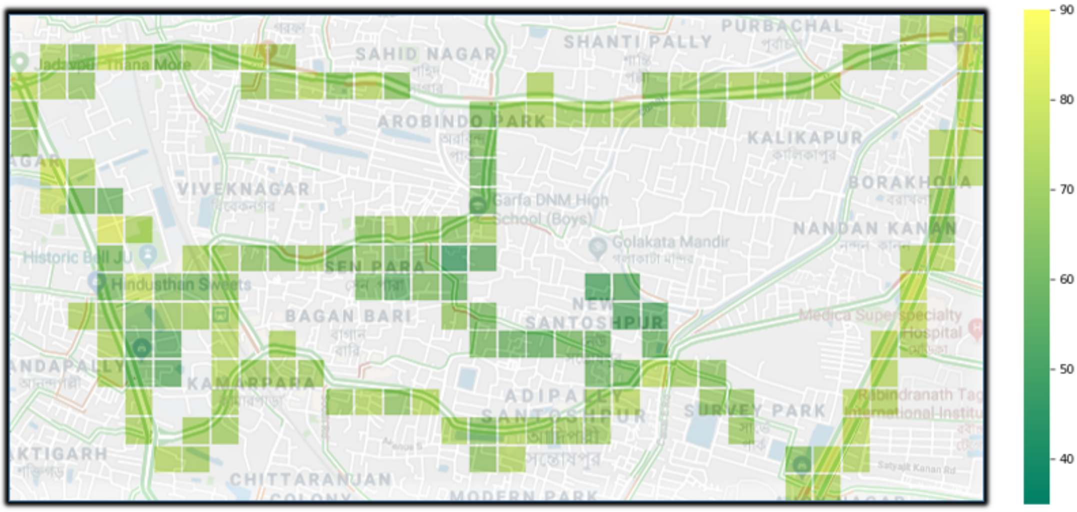

FIGURE 5.6: Noise pollution map generated by Grid-MSM method.

5.4.2 Performance Metrics

We have used four metrics to compare the performance of our proposed methodologies (Grid-MSM, I-Grid-MSM, Grid-MSM-SK, Grid-MSM-OK) with some popular methods (IDW, SK, OK). We have used statistical measures like relative mean absolute error (RMAE) [40] and root mean squared errors (RMSE) [39] [40] to validate spatial prediction accuracy. These statistical methods are defined below:

$$RMAE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_{pred_i} - Y_{ref}|}{Y_{ref}} \tag{5.2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\{Y_{pred_i} - Y_{ref}\}} \tag{5.3}$$

Where, n is the total number of samples in the validation set, Y_{ref} and Y_{pred_i} are the referenced noise and predicted noise values respectively. RMAE and RMSE have used to measure the prediction accuracy of interpolation methods. Higher values of RMAE and RMSE will lead to lower accuracy in the prediction result and vice versa.

5.4.3 Performance comparison with baseline

In this thesis, we have compared our proposed spatial interpolation methods with two standard methods. The first one is inverse distance weighted (IDW), the second method is Simple Kriging (SK). The predicting accuracy of noise levels for Grid-MSM, I-Grid-MSM, Grid-MSM-SK, IDW, and SK are shown in table **??**. The experimental result shows that none of the methods dominate over others in terms of predictive accuracy (RMAE and RMSE) for a day. This is because of data sparsity i.e. the sparsity of sample data may be different for a different session. However, on average I-Grid-MSM has better performance over others and can be a suitable method for generating the noise pollution map.

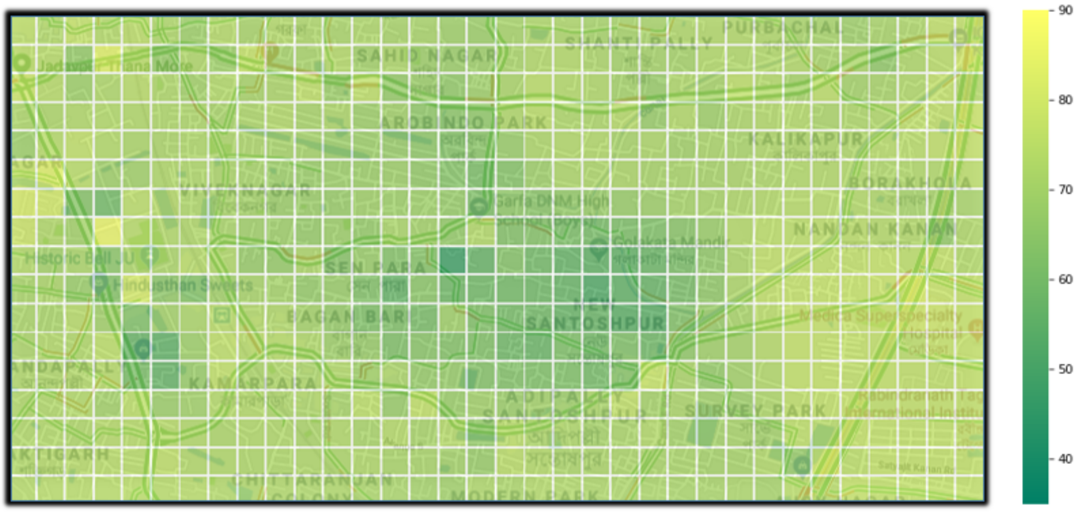

FIGURE 5.7: Noise pollution map generated by I-Grid-MSM method in final iteration.

TABLE 5.1: Performance comparison with baselines

Quality of Models	**Methods**				
	Grid-MSM	*I-Grid-MSM*	*Grid-MSM-SK*	*IDW*	*SK*
RMAE	0.1174	0.0807	0.0994	0.1027	0.1013
RMSE	9.5923	6.9064	7.4322	8.4317	7.7416

RMAE: Relative Mean Absolute Error [40]; RMSE: Root Mean Squared Error [39] [40].
Grid-MSM: Grid based Local Modified Separd's Method; I-Grid-MSM: Iterative Grid based Local Modified Separd's Method; Grid-MSM-SK: Grid based Local Modified Separd's Method Simple Kriging; IDW: Inverse Distance Weighted; SK: Simple Kriging.

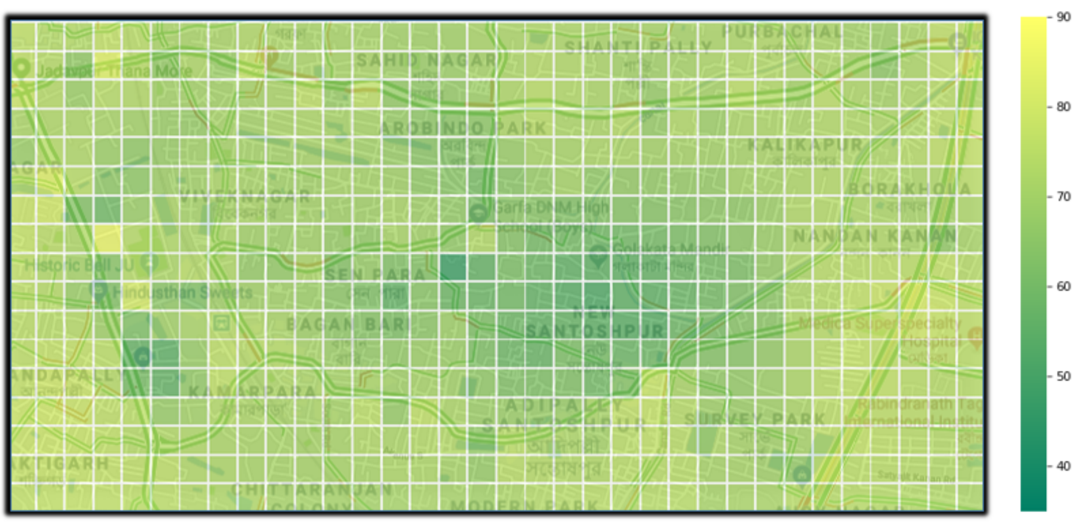

FIGURE 5.8: Noise pollution map generated by Grid-MSM-SK method.

6 Conclusion and Scope for the future works

6.1 Conclusion

The main focus of our work is to create the participatory sensing based dynamic noise pollution map using our proposed methodologies (Grid-MSM, I-Grid-MSM, and Grid-MSM-SK). For this purpose, we have also designed a participatory sensing based framework so that end users can get the noise pollution map and other services depending upon their query.

Our proposed methodologies have been evaluated on different performance metric measurements. We have also compared our methodologies with baseline algorithms like IDW and SK. The experimental results show that I-Grid-MSM is the most suitable approach for a generation participatory sensing based dynamic noise map. I-Grid-MSM infers better results than other methodologies.

6.2 Scope for the future works

I-Grid-MSM method infers better results than other methodologies. In this new methodology, it has both local and global influences for grids surface. It infers well when we fix the time interval. But for analysis of Spatiotemporal sample data, we can not assume temporal as static. As space and time both are continuous in nature, so analyzing a result by making time in static nature may not infer better results. Apart from temporal, contextual information is also a key feature for inferring noise at unmeasured geographical locations. So in the future, we will make a better consideration which will include all these features in an efficient way.

Bibliography

[1] C. Chandramauli, "Census of india 2011: provisional population totals paper 1 of 2011 india series 1, chapter 6," *New Delhi, India: Office of the Registrar General & Census Commissioner*, pp. 438–445, 2011.

[2] Ministry of Environment and Forests, Government of India. No. 14 of 1981, [29/03/1981]-The Air (Prevention and Control of Pollution) Act, 1981, amended 1987. [Last accessed on 2011 Dec 6]. Available from: http://www.moef.nic.in/legis/air/air1.html.

[3] WHO, "Night noise guidelines for europe," http://www.euro.who.int/data/assets/pdffile/0017/43316/E92845.pdf?ua=1, accessed: 2019-0121.

[4] ——, "Burden of disease from environmental noise - quantification of healthy life years lost in europe," https://www.who.int/quantifyingehimpacts/publications/e94888.pdf?ua=1, accessed: 2019-01-21.

[5] Noise: Facts and Figures. World Health Organization. 2011. Available from: http://www.euro.who.int/en/what-we-do/healthtopics/environment-and-health/noise/facts-and-figures. [Last accessed on 2012 Aug 13].

[6] "Environmental noise guidelines for the european region," http://www.euro.who.int/data/assets/pdffile/0008/383921/noise-guidelines-eng.pdf?ua=1, accessed: 2019-01-19.

[7] R. C. Lewis, R. R. Gershon, and R. L. Neitzel, "Estimation of permanent noise-induced hearing loss in an urban setting," Environmental science & technology, vol. 47, no. 12, pp. 6393–6399, 2013.

[8] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," 2006.

[9] Telecom Regulatory Authority of India, IS/ISO 9001-2015 Certified Organisation. Available from: https://main.trai.gov.in/.

[10] Central Pollution Control Board, Ministry of Environment, Forests and Climate Change, Government of India. Available from: http://cpcb.nic.in/Noise_Standards.php. [Last accessed on 2011 Dec 4].

[11] Central Pollution Control Board, Ministry of Environment, Forest and Climate Change, Government of India. Available from: http://cpcb.nic.in/report.php.

[12] J. Dutta, S. Roy, and C. Chowdhury, "Unified framework for iot and smartphone based different smart city related applications," *Microsystem Technologies*, vol. 25, no. 1, pp. 83–96, 2019.

[13] N. Maisonneuve, M. Stevens, M. E. Niessen, and L. Steels, "Noisetube: Measuring and mapping noise pollution with mobile phones," in *Information technologies in environmental engineering.* Springer, 2009, pp. 215–228.

[14] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Earphone: an end-to-end participatory urban noise mapping system," in *Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks.* ACM, 2010, pp. 105–116.

[15] R. Rana, C. T. Chou, N. Bulusu, S. Kanhere, and W. Hu, "Ear-phone: A context-aware noise mapping usingsmart phones," *Pervasive and Mobile Computing*, vol. 17, pp. 1–22, 2015.

[16] M. Becker, S. Caminiti, D. Fiorella, L. Francis, P. Gravino, M. M. Haklay, A. Hotho, V. Loreto, J. Mueller, F. Ricchiuti et al., "Awareness and learning in participatory noise sensing," *PloS one*, vol. 8, no. 12, p. e81638, 2013.

[17] E. Kanjo, "Noisespy: A real-time mobile phone platform for urban noise monitoring and mapping," *Mobile Networks and Applications*, vol. 15, no. 4, pp. 562–574, 2010.

[18] EveryAware Consortium. WideNoise app for Android. Available: http://play.google.com/store/apps/details?id=eu.everyaware.widenoise.android. Accessed 2013 Oct 29.

[19] EveryAware Consortium. WideNoise app for iOs. Available: http://itunes.apple.com/app/id657693514. Accessed 2013 Oct 29.

[20] V. Loreto, M. Haklay, A. Hotho, V. D. Servedio, G. Stumme, J. Theunis, and F. Tria, *Participatory Sensing, Opinions and Collective Awareness.* Springer, 2016.

[21] N. Eagle and A. S. Pentland, "Reality mining: sensing complex social systems," *Personal and ubiquitous computing*, vol. 10, no. 4, pp. 255–268, 2006.

[22] N. Marz and J. Warren, Big Data: *Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co., 2015.

[23] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[24] J. Dutta, F. Gazi, S. Roy, and C. Chowdhury, "Airsense: Opportunistic crowd-sensing based air quality monitoring system for smart city," in *2016 IEEE SENSORS. IEEE*, 2016, pp. 1–3.

[25] J. Dutta, P. Pramanick, and S. Roy, "Noisesense: crowdsourced context aware sensing for real time noise pollution monitoring of the city," in *2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. IEEE, 2017, pp. 1–6.

[26] S. Ghosh, J. Dutta, and S. Roy, "Sensedcity: a participatory sensing based approach," in *Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking*. ACM, 2018, p. 16.

[27] D. Shepard, "A two-dimensional interpolation function for irregularlyspaced data," in *Proceedings of the 1968 23rd ACM national conference*. ACM, 1968, pp. 517–524.

[28] H. Wackernagel, "Multivariate geostatistics: an introduction with applications," in *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts*, vol. 8, no. 33, 1996, p. 363A.

[29] N. S.-N. Lam, "Spatial interpolation methods: a review," *The American Cartographer*, vol. 10, no. 2, pp. 129–150, 1983.

[30] J. Li and A. D. Heap, "A review of spatial interpolation methods for environmental scientists," 2008.

[31] ——, "A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors," *Ecological Informatics*, vol. 6, no. 3-4, pp. 228–241, 2011.

[32] C. Y. Goh, J. Dauwels, N. Mitrovic, M. T. Asif, A. Oran, and P. Jaillet, "Online map-matching based on hidden markov model for real-time traffic sensing applications," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2012, pp. 776–781.

[33] J. W. Hager, L. L. Fry, S. S. Jacks, and D. R. Hill, "Datums, ellipsoids, grids, and grid reference systems," DEFENSE MAPPING AGENCY HYDROGRAPHIC/TOPOGRAPHIC CENTER WASHINGTON DC, Tech. Rep., 1992.

[34] J. W. Hager, J. F. Behensky, and B. W. Drew, "The universal grids: Universal transverse mercator (utm) and universal polar stereographic (ups). edition 1," DEFENSE MAPPING AGENCY HYDROGRAPHIC/TOPOGRAPHIC CENTER WASHINGTON DC, Tech. Rep., 1989.

[35] C. M. Bishop, Pattern recognition and machine learning. springer, 2006.

[36] "Amprobe sm-10 sound level meter," `http://www.amprobe.com/amprobe/`, accessed: 2019-05-11.

[37] "Mextech sl-4012 sound level meter," `https://www.mextech.co.in/sound-level-meter.html`, accessed: 2019-05-11.

[38] "Google Map", `https://www.google.com/maps/@22.4968599,88.3779878,15z/data=!5m1!1e1`, accessed: 2019-05-11.

[39] Chai, Tianfeng, and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature." *Geoscientific model development* 7.3 (2014): 1247-1250.

[40] Shcherbakov, Maxim Vladimirovich, et al, "A survey of forecast error measures." *World Applied Sciences Journal* 24.24 (2013): 171-176.

[41] ArcMap, "How Kriging works", `http://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm`, accessed: 2019-05-11.