GEOSS: an intelligent methodology for identifying site suitability of air sample collection

Kamonasish Mistry

Department of Geography,
Sammilani Mahavidyalaya,
Baghajatin, Kolkata, India
Email: kamonasish79@gmail.com

Biplab Biswas

Department of Geography,
The University of Burdwan,
Burdwan, India
Email: biswas.biplab@gmail.com

Siwen Zhang

School of Economics & Management,
Tong Ji University,
Shanghai, China
Email: zsw5042@126.com

Tao Wu* and Liang Zhou

Shanghai Jiao Tong University School of Medicine,
Shanghai, China
Email: ijista@aliyun.com
Email: wenzhou6@sjtu.edu.cn
*Corresponding authors

Abdelfettah Benchrif

National Centre for Nuclear Energy, Sciences and Technology
(CNESTEN), Morocco
Email: abenchrif@gmail.com

Srimanta Gupta

Department of Environmental Science,
The University of Burdwan,
Burdwan, India
Email: srimantagupta@yahoo.co.in

Abstract: Air pollution (AP) types and levels change with changes in land use land cover (LULC) types. However, there is no such attempt to develop any common methodology or model for optimum sampling which can be correlated between LULC types and changes with the AP level types and changes. A pre-planned, well-calculated geospatial method is needed to evaluate the ambient AP level, type and its variation over different LULC types. '*GEOSS*' (*geospatial estimation of optimum sample site*) has been innovated to identify the optimum AP sampling sites so that it can represent the wide spatial coverage over varied LULC types. Classified satellite images and statistical tools are used to optimize sampling locations. Validation approach based on nearest neighbour analysis (NNA) has justified that GEOSS employed sampling points are systematically distributed and fulfilled all the basic assumptions of the present sampling procedure.

Keywords: geospatial estimation of optimum sample site; GEOSS; geospatial modelling; optimum location; land use land cover; LULC: Kolkata Metropolitan Area; KMA; air pollution level and change; sampling techniques.

Reference to this paper should be made as follows: Mistry. K., Biswas, B., Zhang, S., Wu, T., Zhou, L., Benchrif, A. and Gupta, S. (2020) 'GEOSS: an intelligent methodology for identifying site suitability of air sample collection', *Int. J. Intelligent Systems Technologies and Applications*, Vol. 19, No. 5, pp.421–443.

Biographical notes: Kamonasish Mistry is currently working as an Assistant Professor of Geography in Sammilani Mahavidyalaya under University of Calcutta, West Bengal, India. His current areas of research include air pollution, land use land cover changes and public heath using remote sensing and GIS. Other areas of interest are drainage mapping and soil mapping of Rarh Bengal, India.

Biplab Biswas is a 'Raman Fellow' (USA) and Professor in the Department of Geography, The University of Burdwan, Burdwan, India. He received his PhD from the University of Burdwan in 2008. He is interested to study the environment and public health with particular emphasis on air and water quality. He uses GIS intensively in his teaching and research. He has successfully guided five PhD candidates and two MPhil candidates.

Siwen Zhang received her Masters degree in 2011 from Donghua University, and currently a PhD candidate in School of Economics and Management, Tong Ji University. Her research interest includes machine learning, artificial intelligence and their applications.

Tao Wu currently serving as a Professor in Shanghai Jiaotong University, School of Medicine. His research interests include artificial intelligence and its applications. In recent years, he has published more than 50 research papers and presided more than 20 scientific research projects at provincial level or above.

Liang Zhou received his PhD in Management Science in 2012 from Donghua University, Shanghai, China. Currently, he is serving as an Assistant Professor in Shanghai Jiaotonh University, School of Medicine. His research interests include optimisation computing, simulation, modelling, and artificial intelligence.

Abdelfettah Benchrif is a Research Scientist in the National Centre for Nuclear Energy, Sciences and Techniques (CNESTEN), Morocco. He received his PhD in 2011 from Mohammed V University, Rabat, Morocco. His research interests are focused on the study of urban and regional air quality problems with an emphasis on the size and composition of atmospheric particles. With his colleagues, he uses state-of-the-art analytical techniques in the laboratory and the field to analyse aerosol chemical composition, identify their suitable chemical fingerprints, and quantify their source contributions.

Srimanta Gupta is currently working as an Associate Professor of Environmental Science, The University of Burdwan and has more than 15 years experience in Post-Graduate research and teaching. He has specialisation in surface and groundwater hydrogeochemistry, contaminant transport and geochemical modelling, climate change, air pollution and meteorology, disaster management, remote sensing and GIS. He has published more than 100 research articles in various reputed national and international journals. He has also presented his research in various national and international conferences and has more than 60 conference proceedings. He has worked on several major projects sponsored by Government of India and Government of West Bengal and has supervised 12 PhD students. He is a member of several academic committee, scientific bodies, expert member and reviewer of reputed international and national journals has carried out several societal outreach programmes.

This paper is a revised and expanded version of a paper entitled 'GEOSS: a methodology of sample collection for air pollution over different LULC types' presented at 3rd international Seminar on System Dynamics, Human-Nature Interaction & Societal Discourse, University of Burdwan, Burdwan, India, 18th November 2017.

1 Introduction

Land use land cover (LULC) types play a vital role in level and types of particulate matter (PM) in the air (Zou et al., 2016; Meyer and Turner, 1992; Lambin et al., 2003). Still there is no such methodology or model ever developed for optimum sampling, which can establish the relation between LULC types and changes with the air pollution (AP) level and changes (Figure 1).

However, the enlightenment of AP has known through numerous researches (Mazurek, 2002; Adar et al., 2014; Samoli et al., 2016; Pascal et al., 2013; Kloog et al., 2014; Vidale et al., 2017; Viehmann et al., 2015; Zhang et al., 2015; Monrad et al., 2017; Barman et al., 2010; Ghose et al., 2005; Abelsohn and Stieb, 2011). A pre-planned well-calculated methodology is an ultimate need to evaluate the ambient AP level and its variation over different LULC categories. An attempt has made in this paper to set-up a sampling methodology. geospatial estimation of optimum sampling sites (GEOSS) has been innovated to sample the optimum AP sampling sites over the study area so that the collected AP level data represents the whole study area of varied LULC types (Figure 2).

Many of the studies specially on particulate matters (PM pollution) were made considering AP data as supplied by the state agency(ies) (Ghose et al., 2015; Khan et al., 2017; Anenberg et al., 2016; Khreis et al., 2017) and some studies primarily (field based) estimated the AP level using various sampling techniques (Larkin and Hystad, 2017) and instruments (Table 1).

Figure 1 Schematic model of LULC types, changes and AP (see online version for colours)

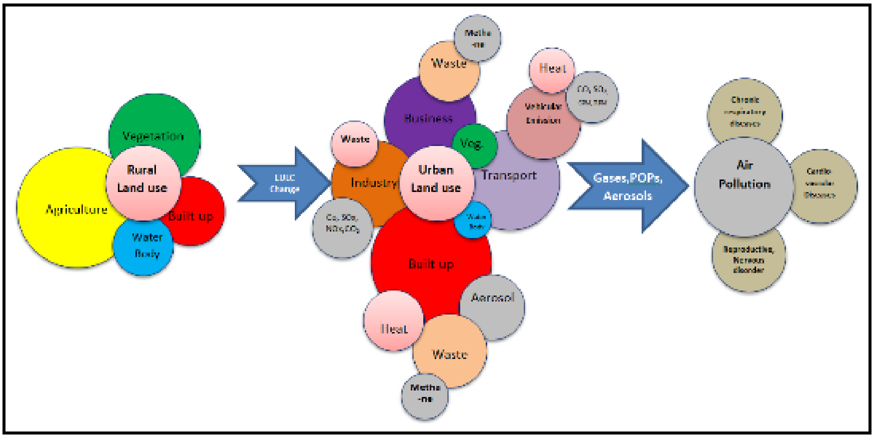


Figure 2 Flow chart of the methodology to determine sample points (see online version for colours)

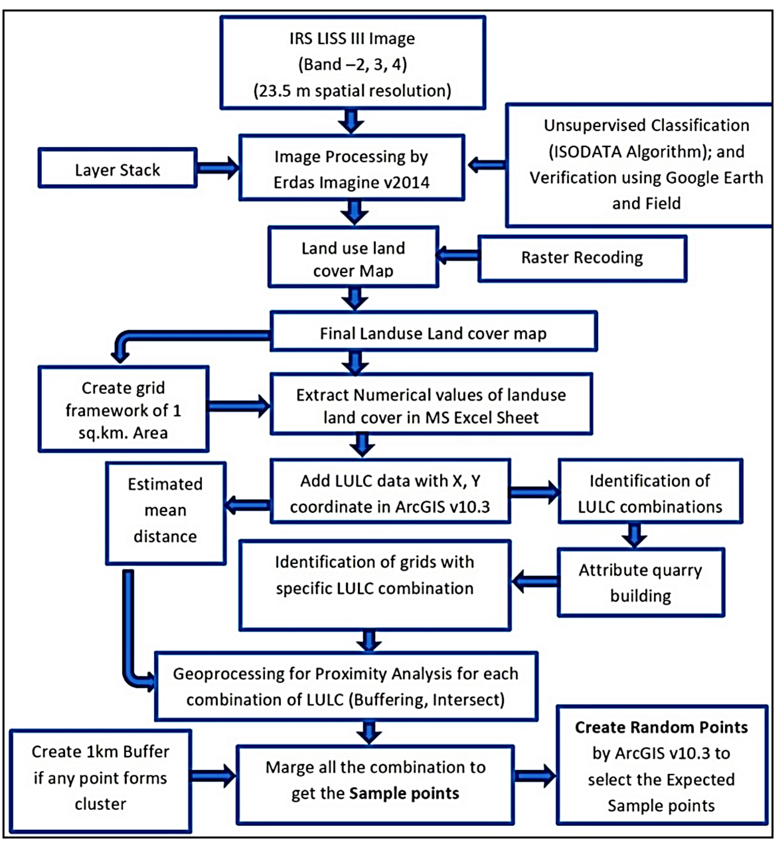
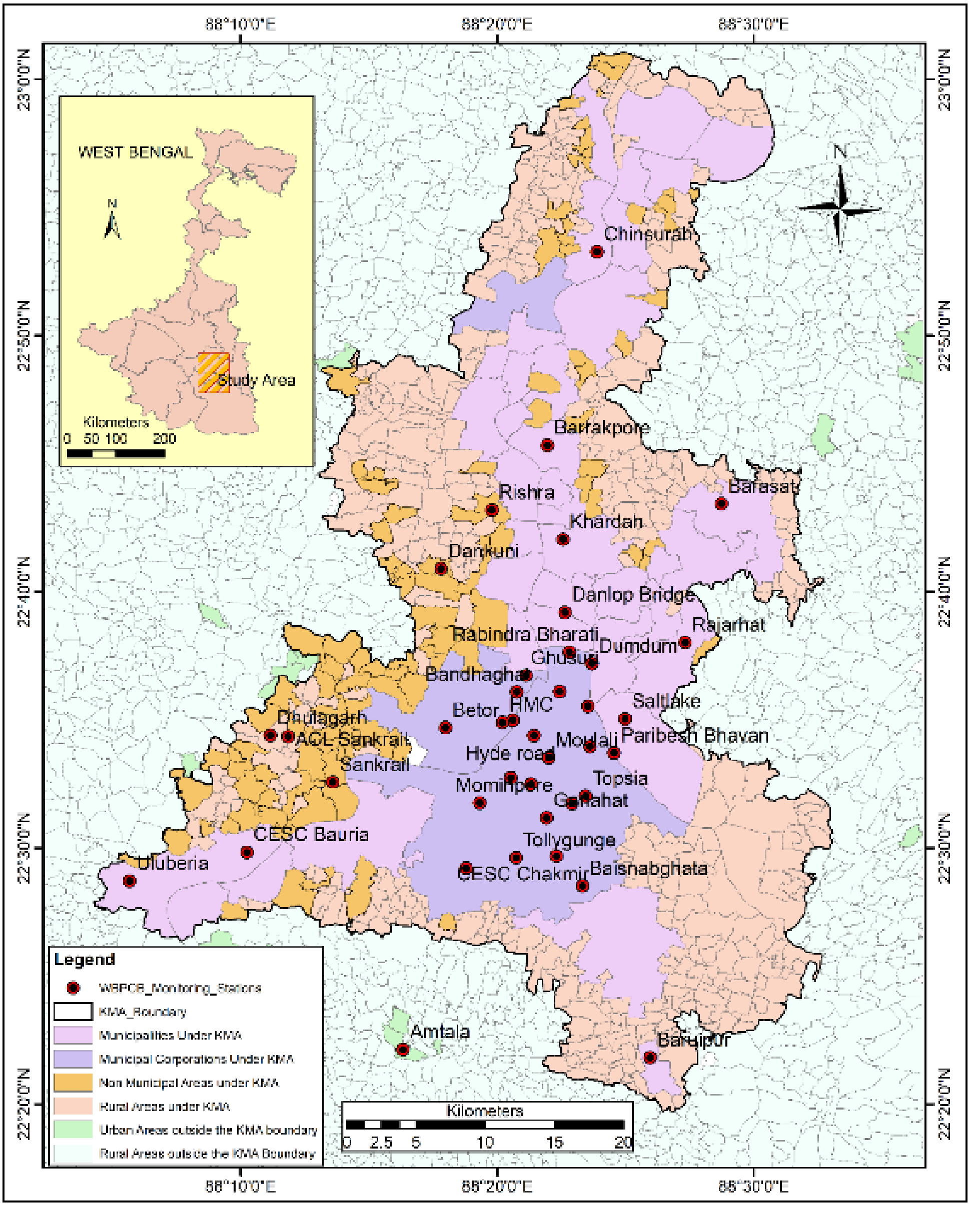


Table 1 Sample collection methods used for collection of PM pollution data from field

Sl no.	*Sample collection method*
1	Dust using vacuum cleaner
2	Dust from AC filters
3	Dust using brush
4	Vacuum cleaner + brush
5	Air sample
6	MiniVol PM2.5 samplers
7	The sampler, made by Science Source, Walboro, ME, USA
8	PM monitors manufactured by the Berkeley Air Monitoring Group
9	Portable constant flow SKC pumps Model 224-PCAR8, SKC, Eighty-Four, PA, USA
10	Harvard impactor
11	Low-volume Leland legacy sampling pumps (SKC, USA)
12	Modified commercial smoke detector; University of California-Berkeley Particle and Temperature Sensor [UCB-PATS]
13	MiniVol™ TAS samplers from
14	APM-550 fine particulate sampler
15	Automated dual channel sampler (HYDRA Dual Sampler, FAI Instruments, Fonte Nuova, Rome, Italy)
16	Harvard personal exposure monitor (PEM)
17	Very quiet PM samplers
18	The personal data RAM-1000AN (pDR-1000)(Thermo Fisher Scientific Inc., Waltham, MA)
19	Personal air sampling pump (SKC AirChek 52)

Almost all of these field based studies have selected the sample sites randomly (Kumar et al., 2008; Guo et al., 2017). On the other hand, the fixed AP measuring stations are highly skewed in distribution. They are located along the major road junctions and industrial units. These fixed stations do not estimate AP level over the various LULC types. As a result, they do not represent the exposure level of all the population residing around each LULC types. For example, in Kolkata and surroundings region, the Central Pollution Control Board (CPCB) and the West Bengal Pollution Control Board (WBPCB) have their fixed (few) stations to collect the ambient AP level data. There are only 39 mobile stations[1] (Figure 3), operated by WBPCB and six fixed stations of CPCB[2] in 2016, whereas the total area and population of the region is 4234 sq. km. and more than 20 million (as per 2011 Census of India) respectively[3]. We can say that one station is measuring exposure for more than 526 thousand populations. Secondly, the stations operate only twice in an entire week which hide temporal exposure of population. Again, the distributions of the AP monitoring stations have clustered within the older urban areas and at the same time, most of them are located on the major road junctions.

Figure 3 Location of WBPCB and CPCB AP monitoring stations (see online version for colours)



The GEOSS has tried to select the sampling points, which have scattered all over the study area and includes all types of LULC categories (Table 2). The existing AP monitoring stations measure PM10 (particulate matter $\leq$ 10 μm in diameter) or more (Gupta et al., 2007; Saygin et al., 2017; Tomaskova et al., 2016), but not the PM5 or PM2.5 (particulate matter < 5 μm or < 2.5 μm).

Table 2 Distribution of sampling points selected with GEOSS

Distribution of points according to administrative area			*Distribution of points according to influence of LULC type*	
Administrative units	*Administrative sub-units*	*Selected sample points*	*LULC type*	*Selected sample points*
Areas inside the KMA boundary	Municipal corporations	12	Water body	7
	Municipalities	18	Urban built-up area	16
	Census town	6	Rural homestead with trees	23
	Villages	10	Agricultural land	17
Areas outside the KMA boundary	Census town	3	Current fallow	4
	Villages	51	Industrial growth centres	7
			Mixed LULC	26
Total selected points	100		Total selected points	100

As per the latest WHO guidelines, PM2.5 directly affects human health both in long-term as well as short-term exposure. We know that Kolkata is one of the oldest city of India and its LULC types are fast changing through the process of urbanisation, urban expansion, industrialisation, (Nath and Acharjee, 2013; Ramachandra et al., 2014) etc. The decadal growth rate of urbanisation of Kolkata Metropolitan Area (KMA) during the period 2001–2011 is 6.87%. However, there is no such attempt to change, update, or re-orient the ambient AP monitoring system to estimate the changing AP level. Moreover, the present AP monitoring stations do not reflect any data over different LULC types. We strongly feel that the existing methods are not suitable for accurate estimation of the ambient AP level for Kolkata and other developing cities. The main objective of this present study is to optimised the sample sites so that it can capture

1 maximum variance in the LULC types
2 maximum spatial distribution of the sampling points to avoid clustering
3 identification of the optimum sampling sites, so that they represent the whole study area of varied LULC types.

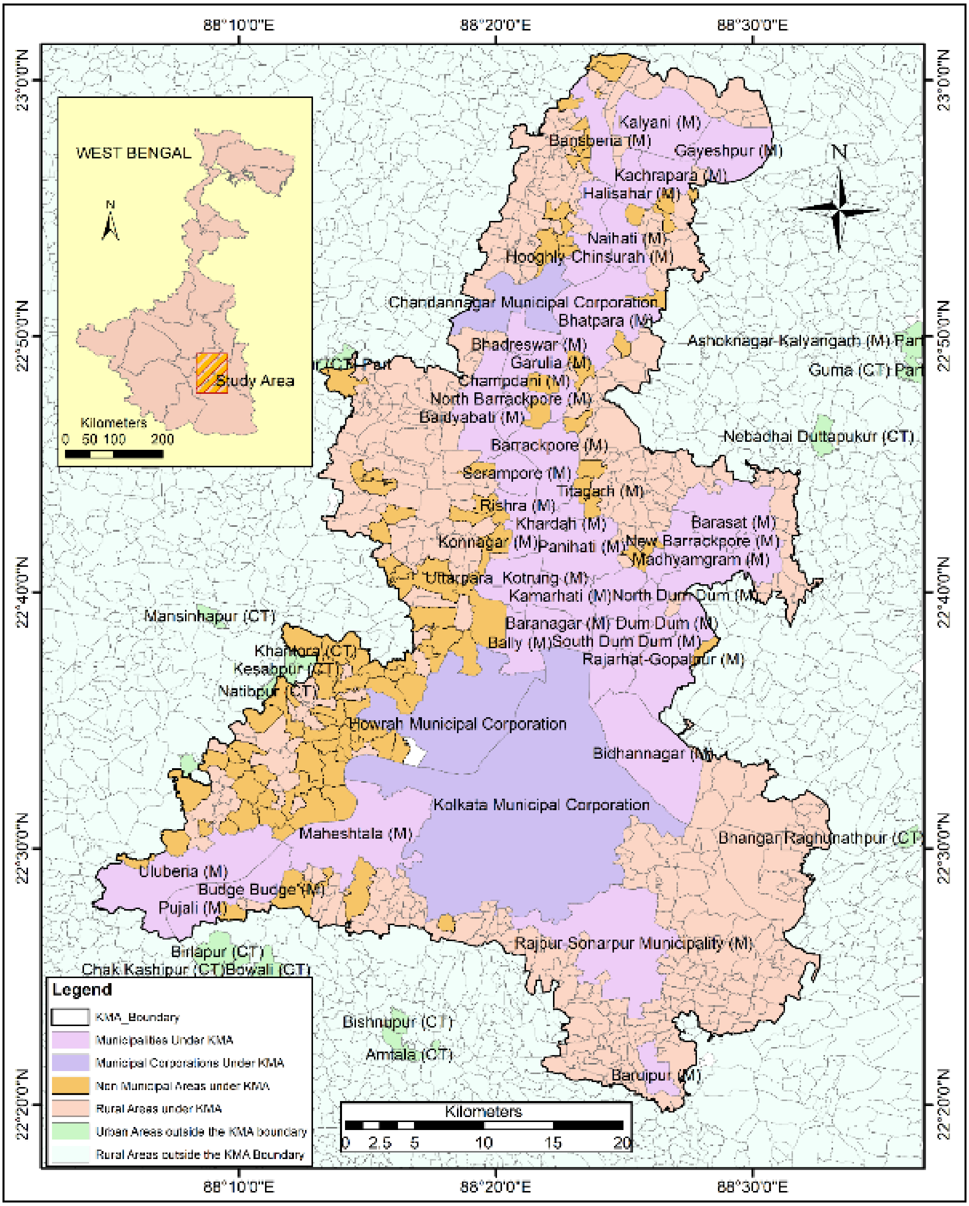
2 Study area

KMA is spread over 1,800 sq. km.[4] (Shaw and Satish, 2007) and comprises of

1 the conurbation area, stretching in a linear manner along the east-west bank of the Hooghly River
2 the rural areas surrounded by the conurbation area.

KMA consists of three municipal corporations (including Kolkata Municipal Corporation), 39 local municipalities, and 107 non-municipal areas; and 442 village local bodies (KMA, 2011)[5]. The suburban areas of KMA include parts of districts of North 24 Parganas, South 24 Parganas, Howrah, Hooghly and Nadia. Latitudinal extent of the study are is from 22° 17’ 23.32”N to 22° 59’ 0.55”N and longitudinally from 88° 4’ 11.59”E to 88° 36’ 58.28”E (Figure 4) serving an estimated total population of about 20 million.

Figure 4 Study area: KMA and surrounding rural area (see online version for colours)



3 Background and feature of GEOSS

The present study area, KMA is one of the oldest city in India. LULC types of KMA and surrounding rural areas are changing through the process of urbanisation, urban expansion, urban renewal, urban sprawl, industrialisation or de-industrialisation (Figure 5, Figure 6, Table 3) (Nath and Acharjee, 2013; Ramachandra et al., 2014; Sen, 2011; Bhatta, 2009) etc.

Figure 5 Google Earth images of Singur industrial site (a) before industrial growth in the year 2004 (b) growth of Tata Industry in the year 2011 (c) after deindustrialisation in the year 2017 (see online version for colours)

(a) (b) (c)

Figure 6 LULC change map of KMA and surrounding during 1990 to 2015 (a) LULC map of the year 1990 (b) LULC category map of the year 2015 (see online version for colours)

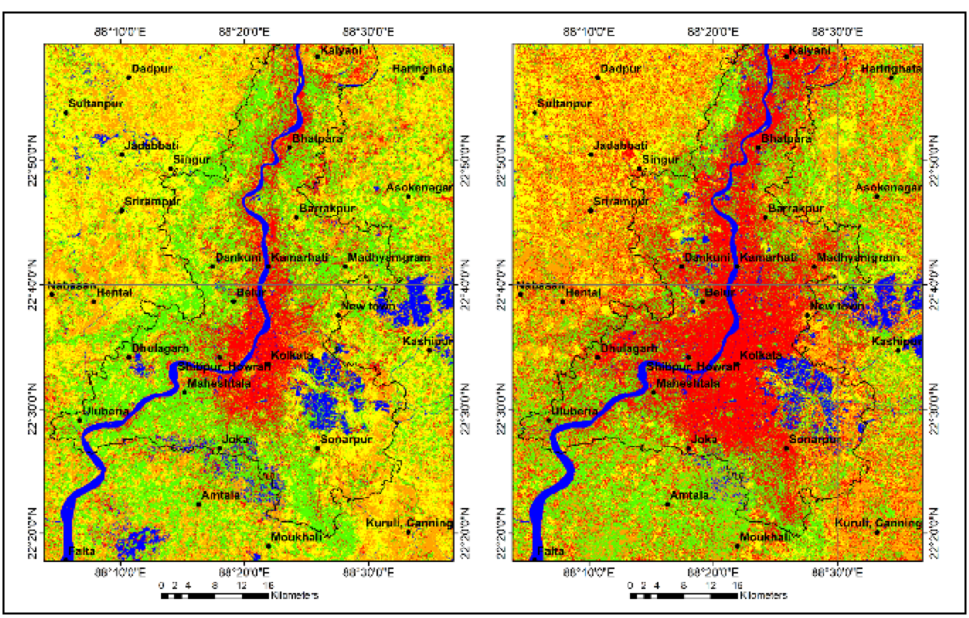


Table 3 LULC change in study area (KMA and surrounding area) during 1990–2015

LULC category	*Area in sq. km. (1990)*	*% of LULC category (1990)*	*Area in sq. km. (2015)*	*% of LULC category (2015)*	*% of LULC change (1990–2015)*
Water body	232.63	*5.49*	219.06	*5.17*	*–5.83*
Trees with homestead	1298.56	*30.67*	1058.93	*25.01*	*–18.45*
Built-up area	615.89	*14.55*	1415.09	*33.42*	*129.76*
Agricultural land	1304.43	*30.81*	949.08	*22.42*	*–27.24*
Current fallow	782.50	*18.48*	591.84	*13.98*	*–24.37*
Total	*4234.00*	*100.00*	*4234.00*	*100.00*	

These processes are also altering sources of AP emission. However, there is no such attempt to change the ambient AP monitoring system to identify the changing sources of AP emission. Kolkata has few fixed and mobile AP monitoring stations mostly for measuring PM10 (particulate matter ≤ 10 μm in diameter). The stations have clustered within the older urban areas at major road junctions and industrial points. However, these are insufficient with respect to total (growing) population and area of KMA. Moreover, the existing stations do not reflect any AP level data from different LULC types. A few studies have collected AP samples from the field, but they have used random sampling (Guo et al., 2017: Gupta et al., 2007) method to collect primary data, which always do not reflect all the LULC types, and most often, they are clustered in distribution.

GEOSS has developed to model the optimum sample sites with the help of geospatial techniques. In this model, geospatial tools have used to estimate LULC types of an area and estimate the optimum sample sites through several geospatial and statistical tools giving emphasis on identification of optimum sample sites over different LULC types to record the role of LULC on AP level and types. The key features of GEOSS are:

a use of geospatial tools for mapping LULC types

b each LULC types and different combinations of LULC types are given equal importance when sampling sites are selected

c a wide spatial coverage to locate the sampling sites

d maintenance of minimum distance between sampling sites to avoid any clustering.

4 Methodology

4.1. Image classification for LULC mapping

The specific area of study (KMA and surrounding area) has selected according to the research interest and prepared an area of interest in the geospatial software Erdas Imagine v2014. The multispectral ortho-rectified image band 2, 3 and 4 (IRS LISS III, Resourcesat-2) of 23.5 metre spatial resolution (Figure 7) have collected from National Remote Sensing Centre, open archive 'Bhuban'. The spectral bands used for image processing, has acquired on 14th December 2015 with 23.50 metre spatial resolution.

Then the integrated image classification method has used to prepare the LULC map of study area (KMA and surrounding area) from satellite image using Erdas Imagine v2014.

Figure 7 Multispectral image of KMA and surroundings (FCC of IRS LISS III image) (see online version for colours)

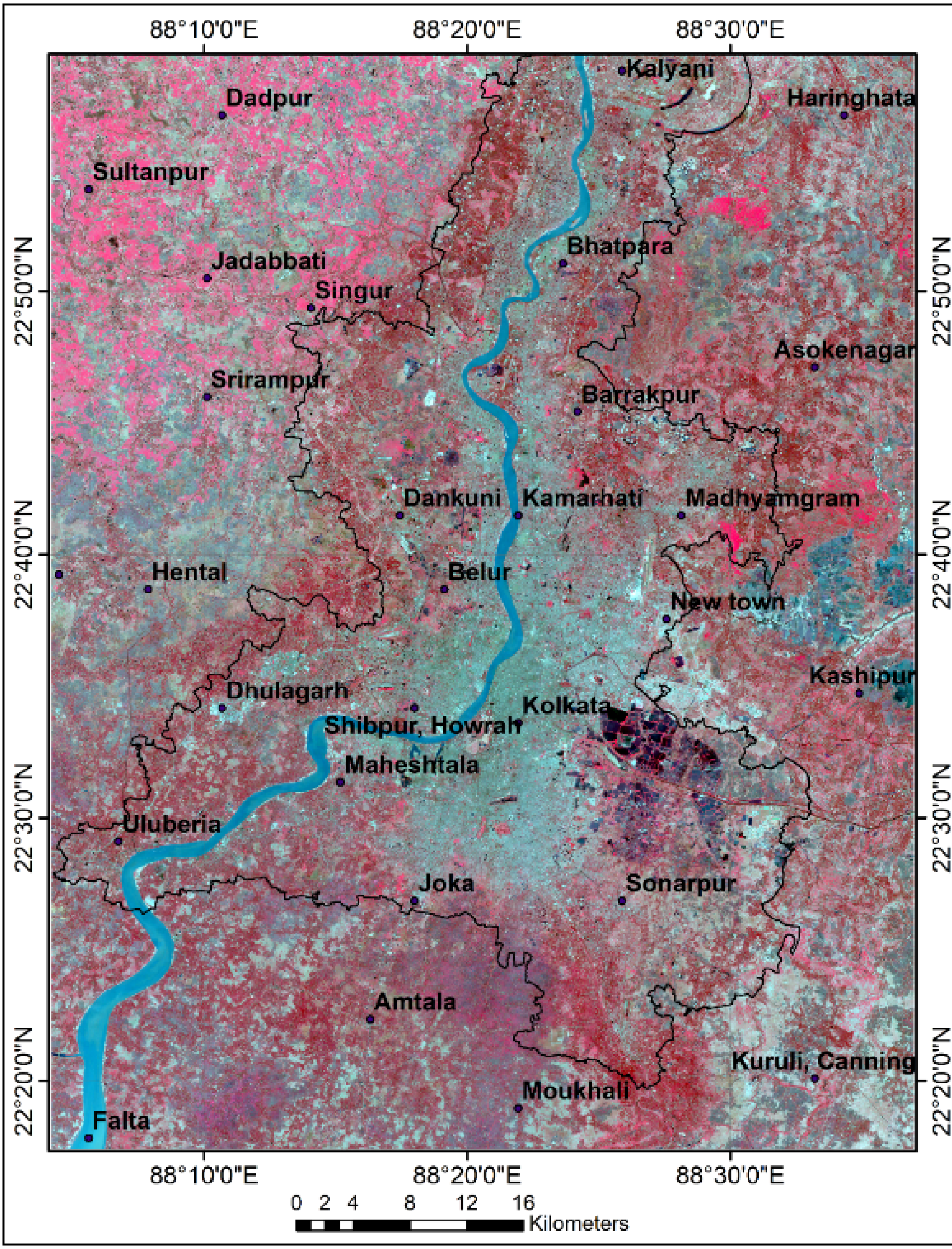
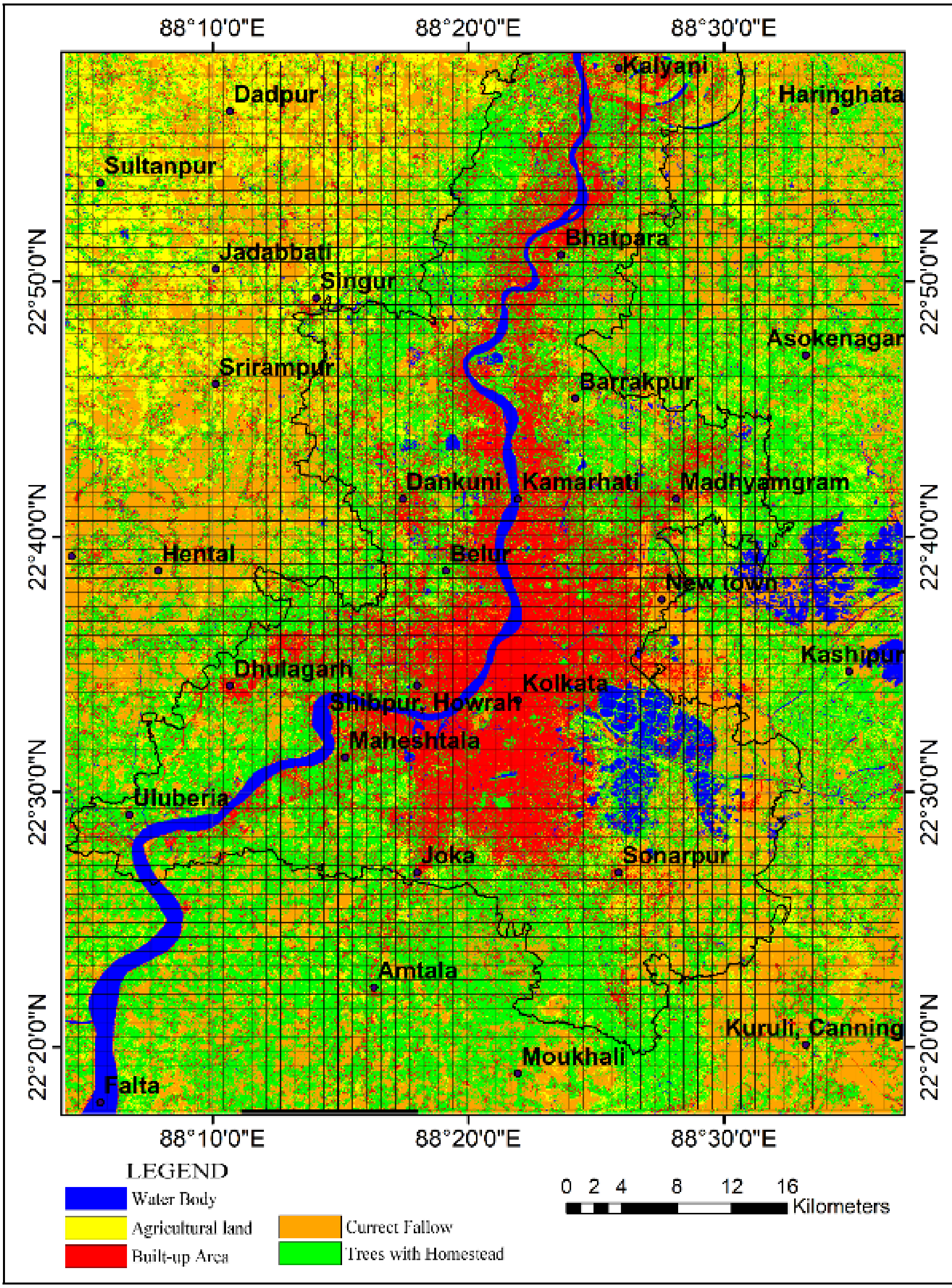


Figure 8 LULC category map with the grid framework of 1 sq. km. (see online version for colours)



Integrated image classification method, combined with both unsupervised (ISODATA) and supervised (Gaussian maximum likelihood) image classification algorithms have used in Erdas Imagine software v2014, to prepare the LULC map of the study area to obtain the better accuracy. Besides, Raster Recode operation has performed to exchange some inaccurate and confused raster signature of the classified LULC category image by assigning actual LULC category signature obtained from ground truth verification and from Google Earth image. This function also used to combine the similar classes by recoding more than one existing class value to the same new class value. After performing the integrated processes of image classification, the final LULC of KMA and surrounding area has prepared. The region has identified five major LULC types, which are:

a water body

b built-up area

c trees with homestead

d agricultural land

e current fallow (Figure 8).

4.2 Creating point location to identify AP sampling sites

To satisfy the criterion of the well-distributed AP sampling sites over the classified LULC map, a systematic method has employed in geospatial environment to generate sampling sites over the map. There are 4234 points with 1 km distance from each other, covering the whole study area. As per our objective, we shall select those points that are significantly fulfilled the sampling objectives and the number of sites (e.g., 100 points) should be measurable. To fulfil the earlier purpose the following steps have adopted.

4.2.1. Creating grid network to extract LULC data

One sq. km. grid framework has generated and superimposed over the LULC map of the study area covering 4234 sq. km. (Figure 7). After that, grid-wise LULC type(s) has been extracted and transferred to a MS Excel sheet with x, y (longitude, latitude) coordinates of the grid centres and observed that each 1 sq. km. grid may comprise (Table 4a):

- single LULC category
- any two LULC categories
- any three LULC categories
- any four LULC categories
- any five LULC categories.

The number of combinations for each set of LULC data has calculated to identify the number of grids/points for each combination. For example, there are five LULC categories in total and if the combination has made for any four categories of LULC, then the expected combinations of any four LULC categories are

$$A+B+C+D,\ A+C+D+E,\ B+C+D+E,\ A+B+D+E \text{ and } A+B+C+E$$

In this way, ten different combinations have identified for any two and three LULC types and five different combinations for any one and four LULC categories and one combination for all the LULC categories. There were 2, 3 and 36 grids containing single

LULC types, any two LULC combinations and any three combinations respectively. There are 752 grids contained any 4 LULC combinations and 3441 for all five LULC categories (Table 4a).

Table 4a LULC combination

Combinations of LULC categories	*No. of combinations*	*Combinations*	*No. of grids*
Any 1	5	A or B or C or D or E	2
Any 2	10	A+B, A+C, A+D, A+E, B+C, B+D, B+E, C+D, C+E and D+E	3
Any 3	10	A+B+C, A+B+D, A+B+E, A+C+D, A+C+E, A+D+E, B+C+D, B+C+E, B+D+E and C+D+E	36
Any 4	5	A+B+C+D , A+C+D+E , B+C+D+E , A+B+D+E and A+B+C+E	752
5	1	A+B+C+D+E	3441
Total	4234		

Note: In LULC combination, A = water body, B = built-up, C = trees with homestead, D = agricultural land and E = current fallow.

4.2.2 Attribute query

As the dataset is very large with number of variables and manually it is difficult to identify the particular combination. Attribute query expression in GIS environment is used to select features that match the selection criteria (Chang et al., 2007). By using attribute query, those particular grids have been identified which contains specific LULC combination. The same selection operation may also be done with python expression. We use attribute quarry in ArcGIS platform because this expression is well known to the users and easy to use.

4.2.3. Proximity analysis/buffer analysis

Buffer has created around each of the identified point, which involved creation of a 'circular' polygon around each point of radius equal to the buffer width (which is a specified distance) by selecting planar buffer method in GIS environment (Figure 9).

The specific distance has calculated between the points in respect of whole study area using the following equation

Expected mean distance:

$$\overline{D}_E = \frac{0.5}{\sqrt{n/A}} \tag{1}$$

where,

n total number of observation points

A total number of study area.

Figure 9 Creating 1 km buffer to identify clusters (see online version for colours)

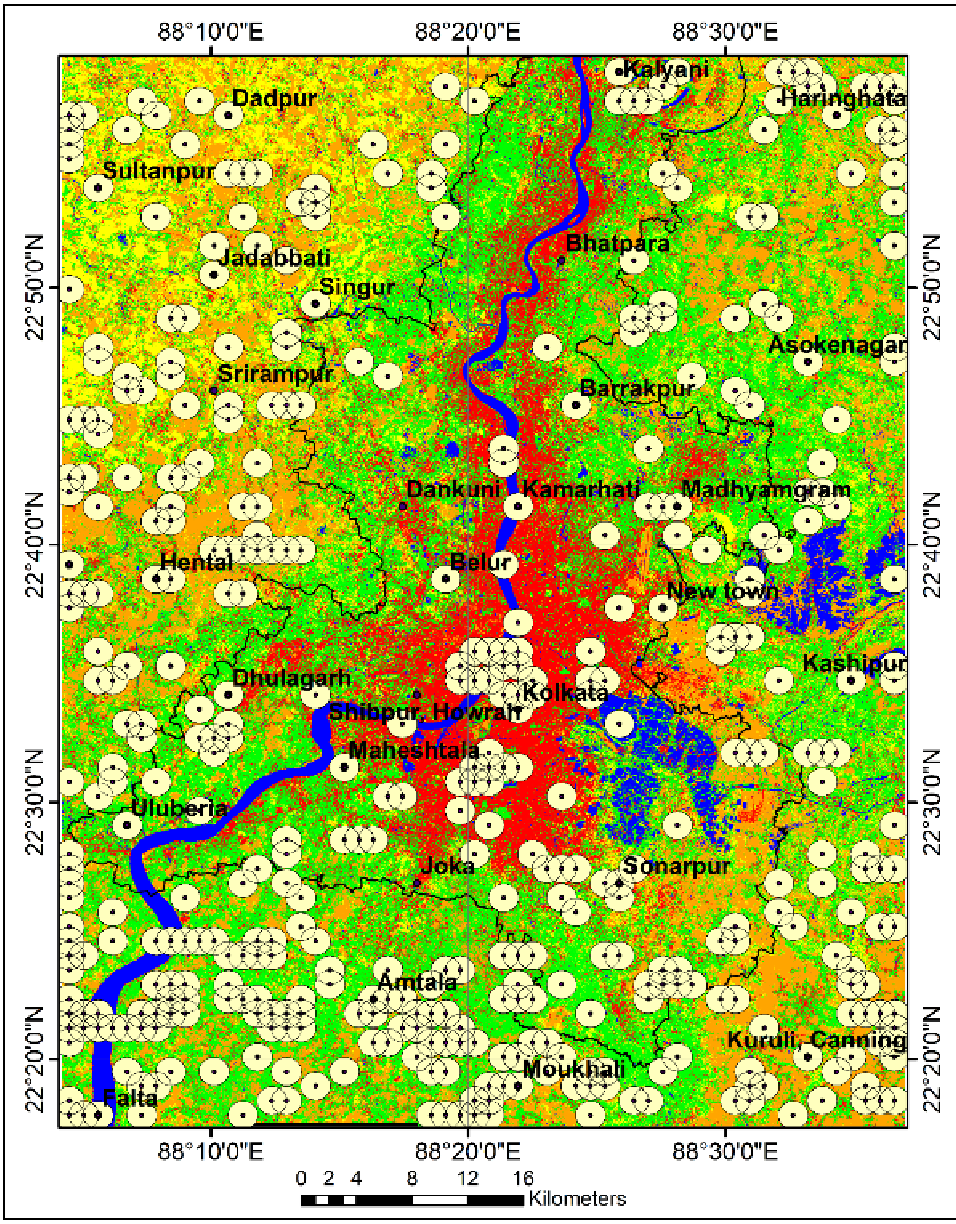


Figure 10 Selected sampling points of the study area (see online version for colours)

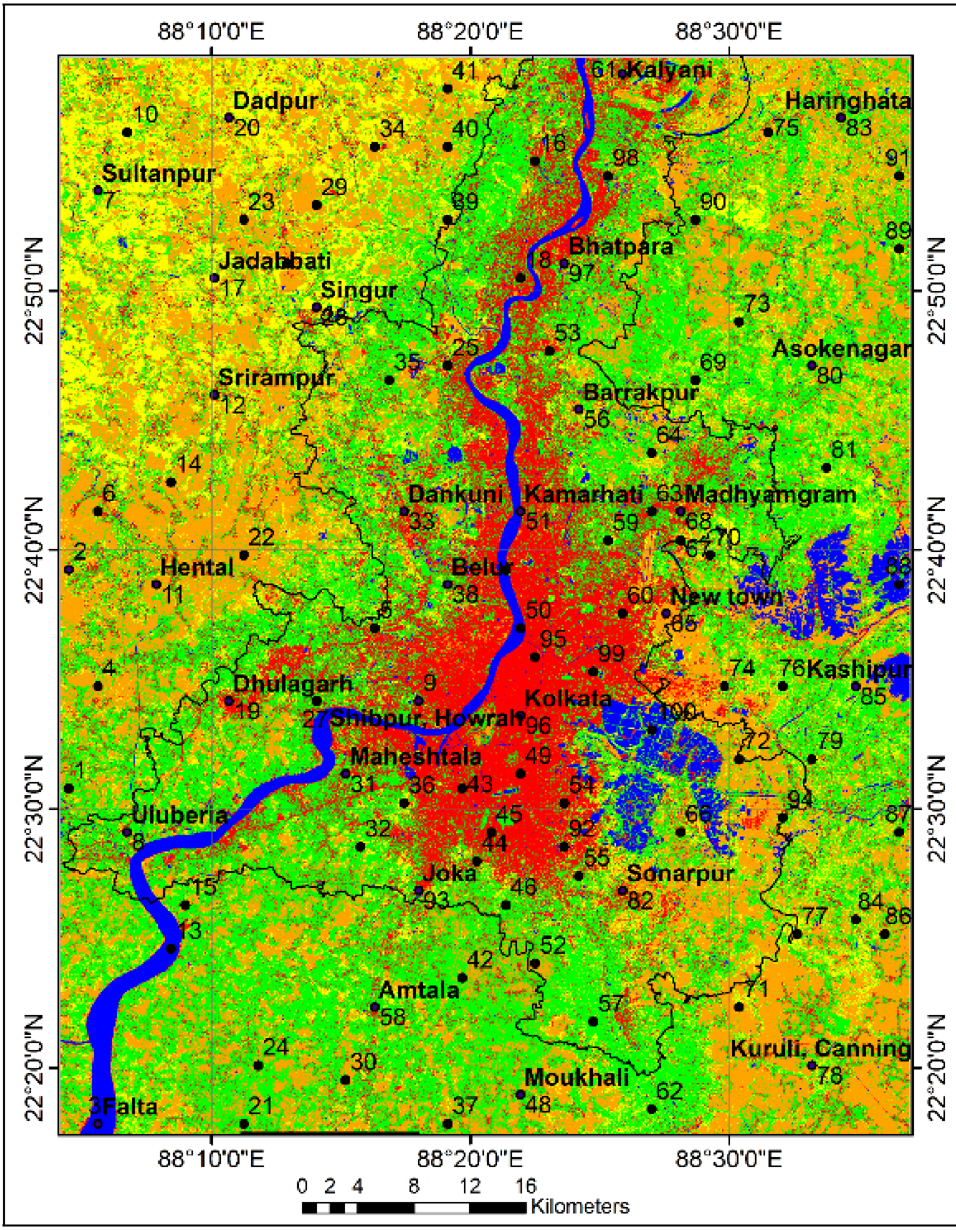


Table 4b Number of selected grids out of 4234 grids

Combinations of LULC categories	*No. of combinations*	*No. of selected grids*
Any 1	5	116
Any 2	10	296
Any 3	10	30
Any 4	5	3
5	1	2
Total	447	

This process has resulted some buffers those have intersected with each other forming clusters. Efforts have been put to extract those points which are free from cluster or which buffers are not intersected each other through location based spatial quarry in GIS environment for separate LULC combination. This has resulted 447 sites with all five different LULC combinations (Table 4b).

After elimination of the clusters from separate LULC combination, the five separate files have merged in a single file. Since 100 sampling sites have to select, again the buffer has generated with 1 km radius so that the minimum distance between the points remain 2 km or more by spatial quarry building. Finally, after elimination of the intersected points, expected number of sampling sites has selected which fulfils the sampling objectives (Figure 10).

Alternatively, 100 points may be selected from 447 points with minimum distance of 2 km by random point selection in Arc-Tool box (ArcGIS v10.3) representing all the LULC combinations. A set of 100 sampling sites (Table 4c) were identified in and around KMA using the GEOSS, the optimum sampling design considering the LULC types and changes to assess PM pollution level, major source region and concentration level and health effects.

5 Results

The present study area has 4234 points with a distance of 1 km. and among them 100 optimum points have identified for AP sampling. The GEOSS has proved useful to identify those 100 optimum locations. The sites represent maximum variations in the LULC; maximum spatial extension without clustering and represented the whole study area which if highly fulfilled the primary assumptions of the sampling objectives. The selected 100 sites contained single LULC type in one site; any two LULC type in one sites; any three LULC types at 12 sites; any four and five LULC types in 67 and 19 sites respectively (Table 4c). Sites are located covering most types of administrative units over urban areas (e.g. municipalities, towns, and cities), rural local bodies, semi urban, urban periphery etc. as well as industrial areas, and congested road junctions.

Table 4c 100 selected sampling points

Combinations of LULC categories	*No. of combinations*	*Selected sample points*
Any 1	5	1
Any 2	10	1
Any 3	10	12
Any 4	5	67
5	1	19
Total points	100	

6 Discussions

Broadly, the LULC types of the study area have re-grouped into urban and rural LULC types and it has observed that 39 sites have selected from the urban and 61 from the rural LULC types. Among 100 selected points, seven sites are located in and around water body and 23 sites belong to urban built-up (seven sites of which are located near the industrial growth centres). 23 sites represent rural homestead with trees; 17 sites are from agricultural LULC types; 4 sites represent current fallow and 37 sites have no distinct LULC types but have a mixed one (Table 2).

In the context of administrative boundaries, there are 46 sites located under KMA boundary from which 12 sites are under municipal corporations, 18 and 6 points located under municipalities and census town respectively. Ten sites are located in rural areas of KMA. Whereas, 54 sites are selected from outside the KMA boundary, only three of which belong to town area (Table 2). The selected points have scattered over the whole study area and any clustering of points has not formed which is clearly enumerated by (Pinder and Witherick, 1972). The randomness index is,

$$R_n = \frac{\overline{D}}{.5\sqrt{\frac{a}{n}}} = 1.428 \quad (2)$$

where,

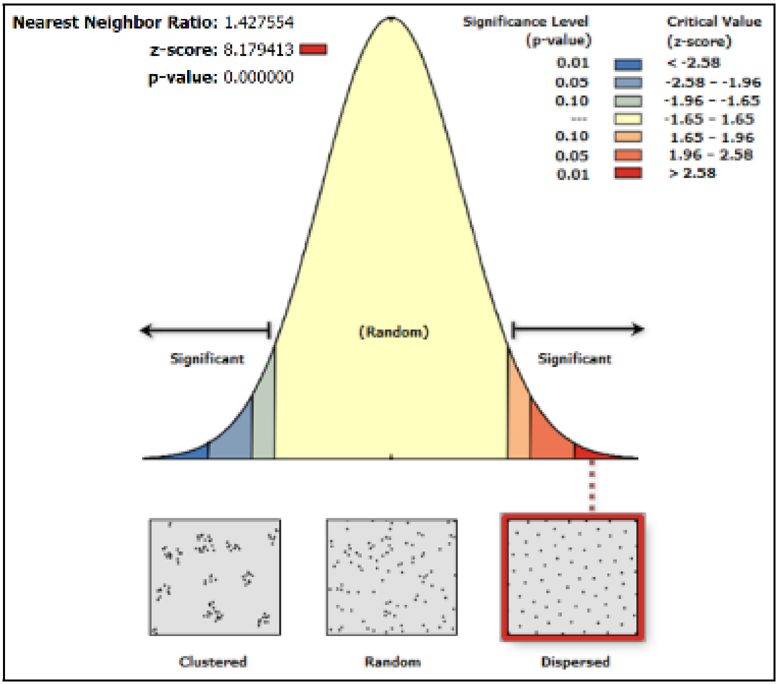
D mean observed nearest neighbour distance (4.47 km.)

a total study area (4234 sq. km.)

n total number of observation points (100 points).

The minimum and the maximum horizontal distance between the points have measured 2 km and 9.7 km respectively, and the observed mean distance between the points has measured 4.47 km. Therefore, it can be said that the final selection of points have fulfilled the sampling objectives. The randomness index (R_n = 1.428) indicates dispersed pattern (Figure 11) of the distribution of sampling points and z-score of 8.179, which is less than 1% likelihood and signifies the dispersed pattern could be result of random chance.

Figure 11 Nearest neighbour result of final sampling points by using GEOSS (see online version for colours)



In comparison with the GEOSS method, 87% of the existing AP monitoring sites (Table 5) of CPCB and WBPCB have clustered in the older urban areas and do not represent different LULC types. The randomness index also indicate the clustered (R_n = 0.739) distribution pattern (Figure 12) of the existing monitoring stations of CPCB and WBPCB. The 45 AP monitoring sites of CPCB and WBPCB collect AP samples but only six of them have fixed to collect hourly data. The other 39 sites collect AP level data with mobile equipment. Out of 45 stations, 41 are located on the main road junctions of urban areas. Only four stations are located near industrial growth points (but they are also situated in the urban areas); and only two points are located in the rural homestead (near urban periphery). Distribution of these sites does not reflect different LULC types.

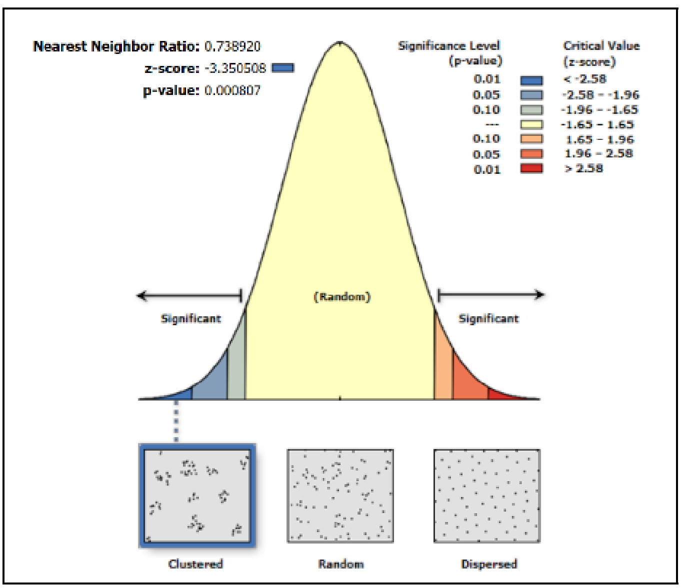
GEOSS, the new and innovative sampling site selection method is more effective for sampling point selection and well distributed having dispersed condition in consideration with varied LULC categories related while existing AP measuring points are clustered and do not reflects all the LULC categories. Therefore, we can conclude that the distribution of points selecting by GEOSS technique is better than the distribution of existing sampling points.

In comparison with previous studies in this regard, most of the studies used random sampling method to collect the samples. However, there is lack of studies to correlate the LULC categories and changes with the change of aerosols as air pollutant and AP levels. Most often, the stations have clustered in some major road junctions on urban areas and industrial centres. The main objective of GEOSS, how far the variations and alterations of LULC responsible for the level and changes of AP do not prioritised in the previous studies. This is highly maintained in GEOSS sampling method.

Table 5 Comparison of the distribution of the points selected by using GEOSS and the existing AP monitoring stations of CPCB and WBPCB according to influence of LULC type

LULC type	*Selected sample points with GEOSS*		*Existing AP monitoring stations*	
	No. of points	*% distribution*	*No. of points*	*% distribution*
Water body	7	7	0	0
Urban built-up area	16	16	39	86.7
Rural homestead with trees	23	23	2	4.4
Agricultural land	17	17	0	0
Current fallow	4	4	0	0
Industrial growth centres	7	7	4	8.9
Mixed LULC	26	26	0	0
Total points	100	100	45	100

Figure 12 Nearest neighbour result of the exiting AP sampling points of CPCB and WBPCB (see online version for colours)



7 Conclusions

GEOSS, the innovative sampling technique has designed to identify the optimum sampling sites over the large study area so that the collected AP level data represent the whole study area of varied LULC types by using geospatial and statistical techniques. While all the prior methods have referred for random sampling, and did not consider LULC categories, but GEOSS has tried to overcome the major issues of random sampling and gave an emphasis to collect air samples from all the LULC types from wide spatial coverage. A set of 100 sampling sites were identified in and around KMA and surrounding rural areas using the GEOSS, considering all major LULC types and changes to assess major source regions, pollution level and concentration level of particulate matter and their health effects. The sample sites were optimised such it covers the spatial coverage of the whole study area as well as scattered over different LULC categories.

After the final selection of the sampling points, a high volume air sampler may be deploy for the rapid collection and analysis of a wide range of airborne aerosols within prerequisite condition of the sampling area and some predefined sampling protocol was maintained. The collected AP level sample data will correlate with the MODIS data for regional distribution of AP level.

8 Acronyms and abbreviations

AP	Air pollution
LULC	Land use land cover
GEOSS	Geospatial estimation of optimum sample site
PM	Particulate matters
CPCB	Central Pollution Control Board
WBPCB	West Bengal Pollution Control Board
KMA	Kolkata Metropolitan Area

References

Abelsohn, A. and Stieb, D.M. (2011) 'Health effects of outdoor air pollution: approach to counseling patients using the air quality health index', *Canadian Family Physician*, Vol. 57, No. 8, pp.881–887.

Adar, S.D., Filigrana, P.A., Clements, N. and Peel, J.L. (2014) 'Ambient coarse particulate matter and human health: a systematic review and meta-analysis', *Curr. Environ. Health Rep.*, Vol. 1, No. 3, pp.258–274.

Anenberg, S.C., Belova, A., Brandt, J., Fann, N., Greco, S., Guttikunda, S., Heroux, M.E., Hurley, F., Krzyzanowski, M., Medina, S., Miller, B., Pandey, K., Roos, J. and Van Dingenen, R. (2016) 'Survey of Ambient air pollution health risk assessment tools', *Risk Anal*, Vol. 36, No. 9, pp.1718–36.

Barman, S.C., Kumar, N., Singh, R., Kisku, G. C., Khan, A.H., Kidwai, M.M., Murthy, R.C., Negi, M.P.S., Pandey, P., Verma, A. K., Jain, G. and Bhargava, S.K. (2010) 'Assessment of urban air pollution and it's probable health impact', *Journal of Environmental Biology*, Vol. 31, No. 6, pp.913–920.

Bhatta, B. (2009) 'Analysis of urban growth pattern using remote sensing and GIS: a case study of Kolkata, India', *International Journal of Remote Sensing*, Vol. 30, No. 18, pp.4733–4746.

Chang, R., Lai, L.B., Su, W.D., Wang, J.C. and Kouh, J.S. (2007) 'Intrusion detection by backpropagation neural networks with sample-query and attribute-query', *International Journal of Computational Intelligence Research*, Vol. 3, No. 1, pp.6–10.

Ghose, M.K., Paul, R. and Banerjee, R.K. (2005) 'Assessment of the status of urban air pollution and its impact on human health in the City of Kolkata', *Environ. Monit. Assess.*, Vol. 108, Nos. 1–3, pp.151–67.

Ghose, S.D., Paul, R. and Banerjee, R.K. (2005) 'Assessment of the status of urban air pollution and its impact on human health in the City of Kolkata', *Environmental Monitoring and Assessment*, Vol. 105, Nos. 1–3, pp.151–167.

Guo, L.Q., Zhang, Q., Zhao, D. D., Wang, L. L., Chen, Y., Mi, B. B., Dang, S.N. and Yan, H. (2017) 'Relationship between air pollution exposure during pregnancy and birth weight of term singleton live-birth newborns', *Zhonghua Liu Xing Bing Xue Za Zhi*, Vol. 38, No. 1, pp.1399–1403.

Gupta, A.K., Karar, K. and Srivastava, A. (2007) 'Chemical mass balance source apportionment of PM10 and TSP in residential and industrial sites of an urban region of Kolkata, India', *Journal of Hazardous Materials*, Vol. 142, Nos. 1–2, pp.279–287.

Khan, M.U. Besis, A., Li, J., Zhang, G. and Malik, R.N. (2017) 'New insight into the distribution pattern, levels, and risk diagnosis of FRs in indoor and outdoor air at low- and high-altitude zones of Pakistan: implications for sources and exposure', *Chemosphere*, Vol. 184, No. 10, pp.1372–1387.

Khreis, H., Kelly, C., Tate, J., Parslow, R., Lucas, K. and Nieuwenhuijsen, M. (2017) 'Exposure to traffic-related air pollution and risk of development of childhood asthma: a systematic review and meta-analysis', *Environ. Int.*, Vol. 100, No. 3, pp.1–31.

Kloog, I., Nordio, F., Zanobetti, A., Coull, B.A., Koutrakis, P. and Schwartz, J.D. (2014) 'Short term effects of particle exposure on hospital admissions in the Mid-Atlantic states: a population estimate', *PLoS One*, Vol. 9, No. 2, p.e88578.

Kumar, N., Chu, A. and Foster, A. (2008) 'Remote sensing of ambient particles in Delhi and its environs: estimation and validation', *Int. J. Remote Sens.*, Vol. 29, No. 12, pp.3383–3405.

Lambin, E.F., Geist, H.J. and Lepers, E. (2003) 'Dynamics of land-use and land-cover change in tropical regions', *Annual Review of Environment and Resources*, Vol. 28, No. 1, pp.205–241.

Larkin, A. and Hystad, P. (2017) 'Towards personal exposures: how technology is changing air pollution and health research', *Curr. Environ. Health Rep*., Vol. 4, No. 4, pp.463–471.

Mazurek, M.A (2002) 'Molecular identification of organic compounds in atmospheric complex mixtures and relationship to atmospheric chemistry and sources', *Environmental Health Perspectives*, Vol. 110, No. 6, pp.996–1003.

Meyer, W.B. and B.L. Turner (1992) 'human population growth and global land-use/cover change', *Annual Review of Ecology and Systematics*, Vol. 23, No. 1, pp.39–61.

Monrad, M., Sajadieh, A., Christensen, J.S., Ketzel, M., Raaschou-Nielsen, O., Tjonneland, A., Overvad, K., Loft, S. and Sorensen, M. (2017) 'Long-term exposure to traffic-related air pollution and risk of incident atrial fibrillation: a cohort study', *Environ. Health Perspect*., Vol. 125, No. 3, pp.422–427.

Nath, B. and Acharjee, S. (2013) 'Urban municipal growth and landuse change monitoring using high resolution satellite imageries and secondary data', *Studies in Surveying and Mapping Science (SSMS)*, Vol. 1, No. 3, pp.43–54.

Pascal, M., Corso, M., Chanel, O., Declercq, C., Badaloni, C., Cesaroni, G., Henschel, S., Meister, K., Haluza, D., Martin-Olmedo, P., Medina, S. and Aphekom, G. (2013) 'Assessing the public health impacts of urban air pollution in 25 European cities: results of the Aphekom project', *Sci. Total Environ.*, Vol. 449, No. 1, pp.390–400.

Pinder, D.A. and Witherick, M.E. (1972) 'The principles, practice and pitfalls of nearest-neighbour analysis', *Geography*, Vol. 57, No. 4, pp.277–288.

Ramachandra, T.V., Aithal, B.H. and Sowmyashree, M.V. (2014) 'Urban structure in Kolkata: metrics and modelling through geo-informatics', *Applied Geomatics*, Vol. 6, No. 4, pp.229–244.

Samoli, E., Atkinson, R.W., Analitis, A., Fuller, G.W., Green, D.C., Mudway, I., Anderson, H.R. and Kelly, F.J. (2016) 'Associations of short-term exposure to traffic-related air pollution with cardiovascular and respiratory hospital admissions in London, UK', *Occup. Environ. Med.*, Vol. 73, No. 5, pp.300–307.

Saygin, M., Gonca, T., Ozturk, O., Has, M., Caliskan, S., Has, Z.G. and Akkaya, A. (2017) 'To investigate the effects of air pollution (PM10 and SO2) on the respiratory diseases asthma and chronic obstructive pulmonary disease', *Turk. Thorac. J.*, Vol. 18, No. 2, pp.33–39.

Sen, S. (2011) 'Effect of urban sprawl on human habitation in urban fringe and peri-urban areas in Kolkata Metropolitan Area', *Institute of Town Planners, India Journal*, Vol. 8, No. 4, pp.58–66.

Shaw, A. and Satish, M.K. (2007) 'Metropolitan restructuring in post liberalized India: Separating the global and the local', *Cities*, Vol. 24, No. 2, pp.148–163.

Tomaskova, H., Tomasek, I., Slachtova, H., Polaufova, P., Splichalova, A., Michalik, J., Feltl, D., Lux, J. and Marsova, M. (2016) 'PM10 air pollution and acute hospital admissions for cardiovascular and respiratory causes in Ostrava', *Cent. Eur. J. Public Health*, Vol. 24, Suppl: pp.S33–S39.

Vidale, S., Arnaboldi, M., Bosio, V., Corrado, G., Guidotti, M., Sterzi, R. and Campana, C. (2017) 'Short-term air pollution exposure and cardiovascular events: a 10-year study in the urban area of Como, Italy', *Int. J. Cardiol.*, Vol. 248, No. 12, pp.389–393.

Viehmann, A., Hertel, S., Fuks, K., Eisele, L., Moebus, S., Mohlenkamp, S., Nonnemacher, M., Jakobs, H., Erbel, R., Jockel, K.H., Hoffmann, B. and Heinz Nixdorf Recall Investigator, G. (2015) 'Long-term residential exposure to urban air pollution, and repeated measures of systemic blood markers of inflammation and coagulation', *Occup. Environ. Med.*, Vol. 72, No. 9, pp.656–63.

Zhang, Z., Laden, F., Forman, J.P. and Hart, J.E. (2016) 'Long-term exposure to particulate matter and self-reported hypertension: a prospective analysis in the nurses' health study', *Environ. Health Perspect.*, Vol. 124, No. 9, pp.1414–20.

Zou, B., Xu, S., Sternberg, T. and Fang, X. (2016) 'Effect of land use and cover change on air quality in urban sprawl', *Sustainability*, Vol. 8, No. 7, p.677.

Notes

1 WBPCB official website www.wbpcb.gov.in

2 CPCB data has not considered for the pollution level calculation by the three WBPCB. The WBPCB says that "Air quality data of automatic stations are not validated data"

4 Census of India, 2011

5 Kolkata Metropolitan Development Authority website https://kmdaonline.org/ Census of India, 2011