INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

Civil Engineering Department

17th Urban Mobility India Conference & Expo 2024

Modeling on-road air pollution using mobile monitoring: A case study of Delhi

Vikram Singh and Amit Agarwal

Presenter:

Vikram Singh

Presentation outline

Introduction

Air pollution

Need of the study

Mobile monitoring

Results

Model development

Model validation

References

Introduction

- Air pollution is one of the **world's leading risk factors** for death, attributed to **millions of deaths** each year (HEI Report 2021).
- Each year, **more people die** from **air pollution-related diseases** than from **road traffic injuries** or **malaria** (State of Global Air/2019).

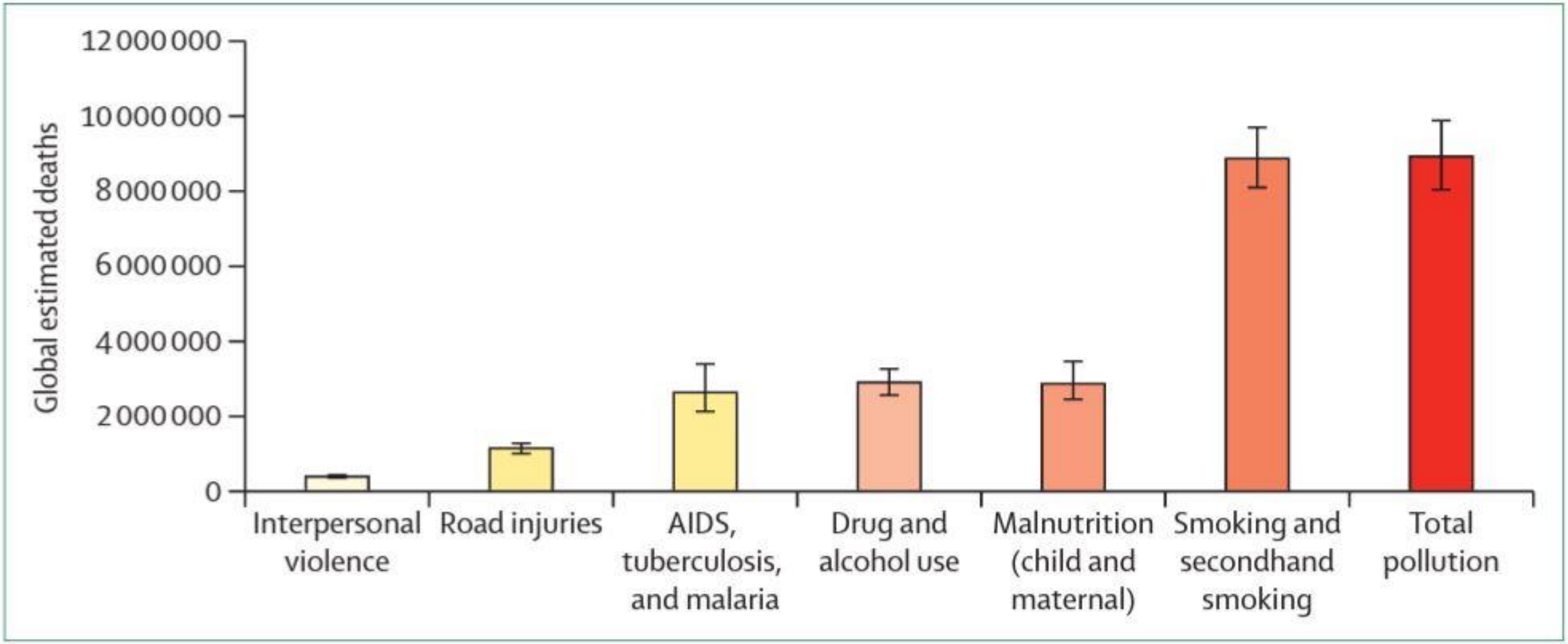
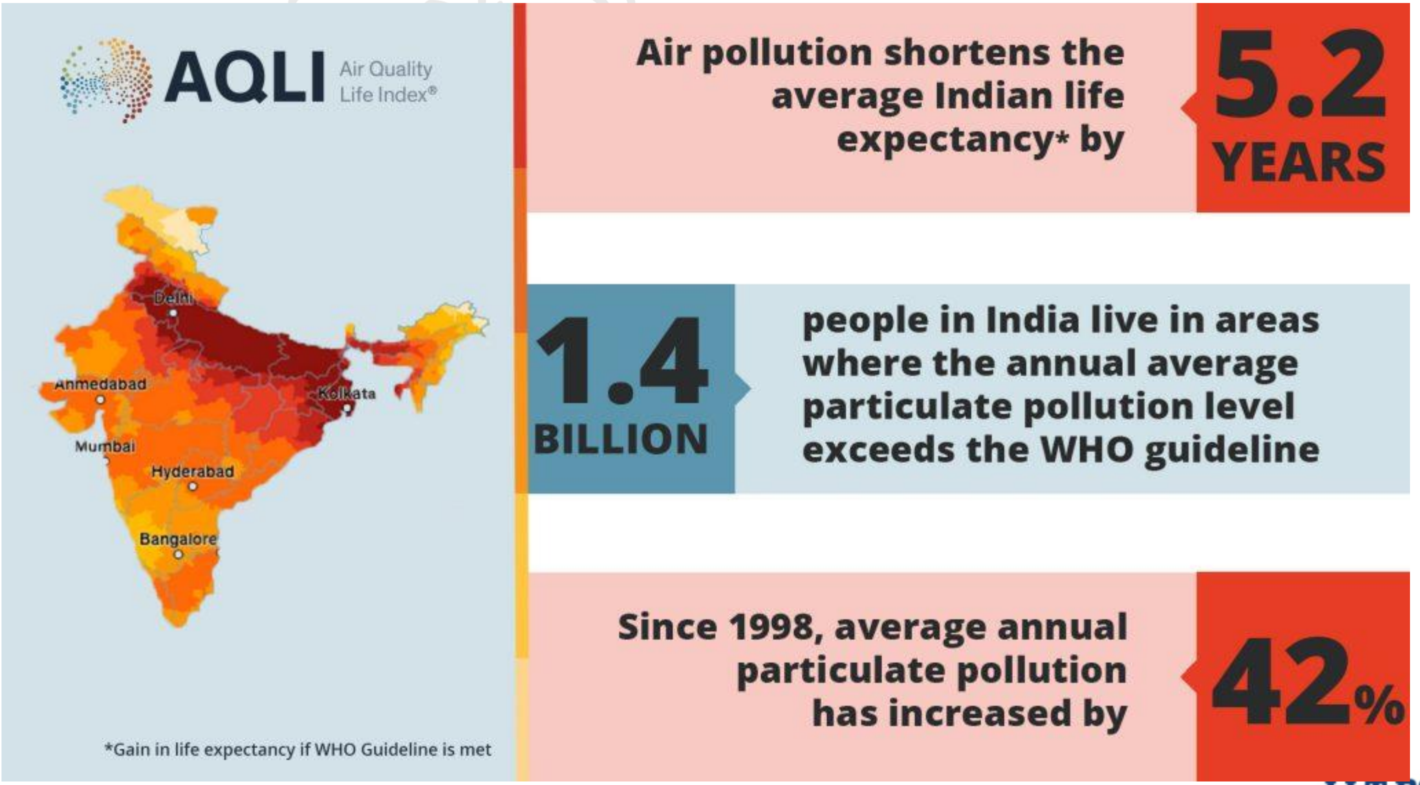


Figure: Major risk factor for cause of global estimated deaths

Source: (Fuller et al. 2022)

Indian scenario

- **India** is **second most polluted country** in the World (Greenstone and Fan, 2020).
- In India, **1.7 million deaths** were due to air pollution in **2019** i.e., **18%** of the **total deaths** in the country (The Lancet Planetary Health Study, 2021).



Change in $PM_{2.5}$ exposure in last decade

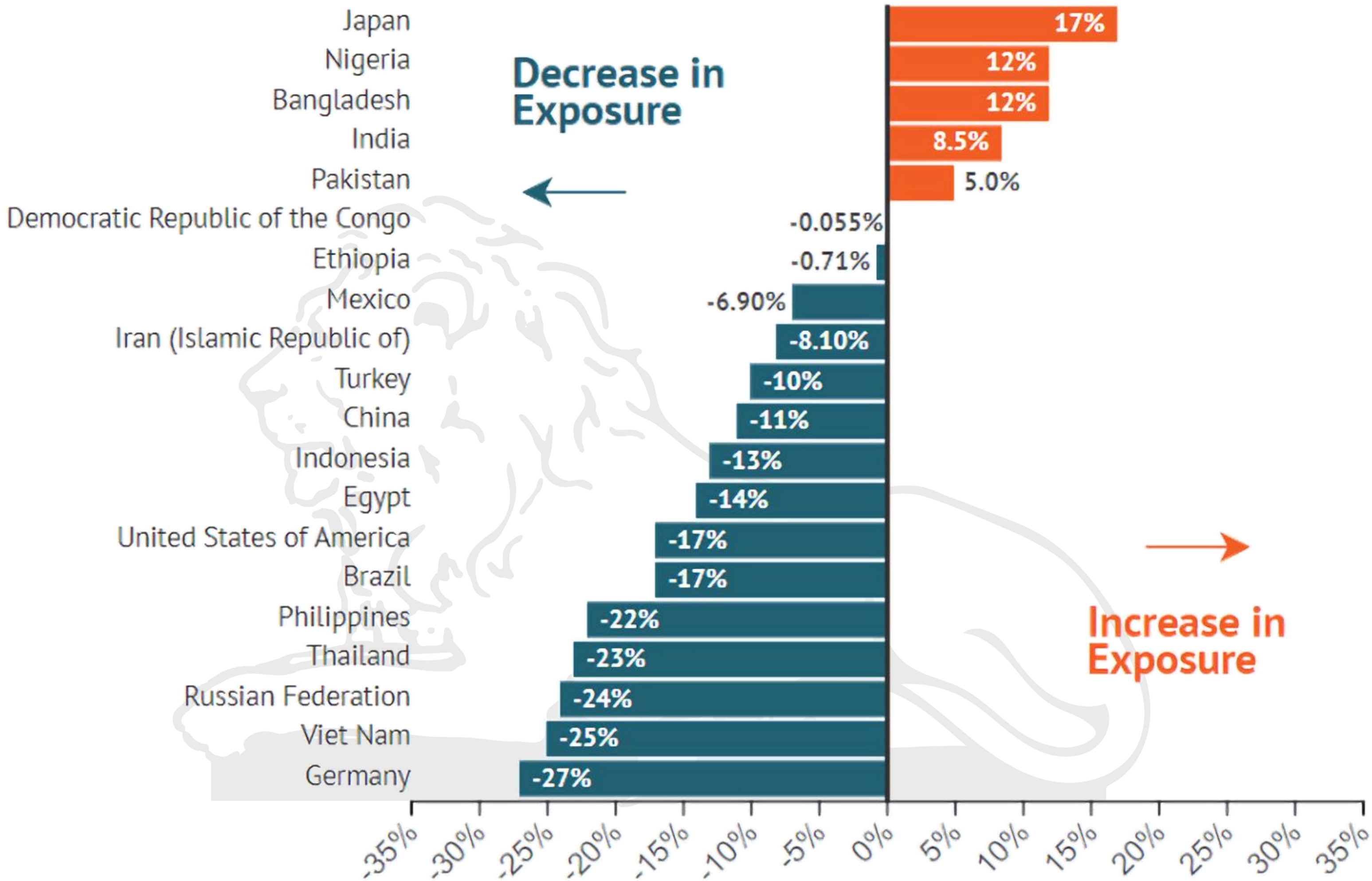
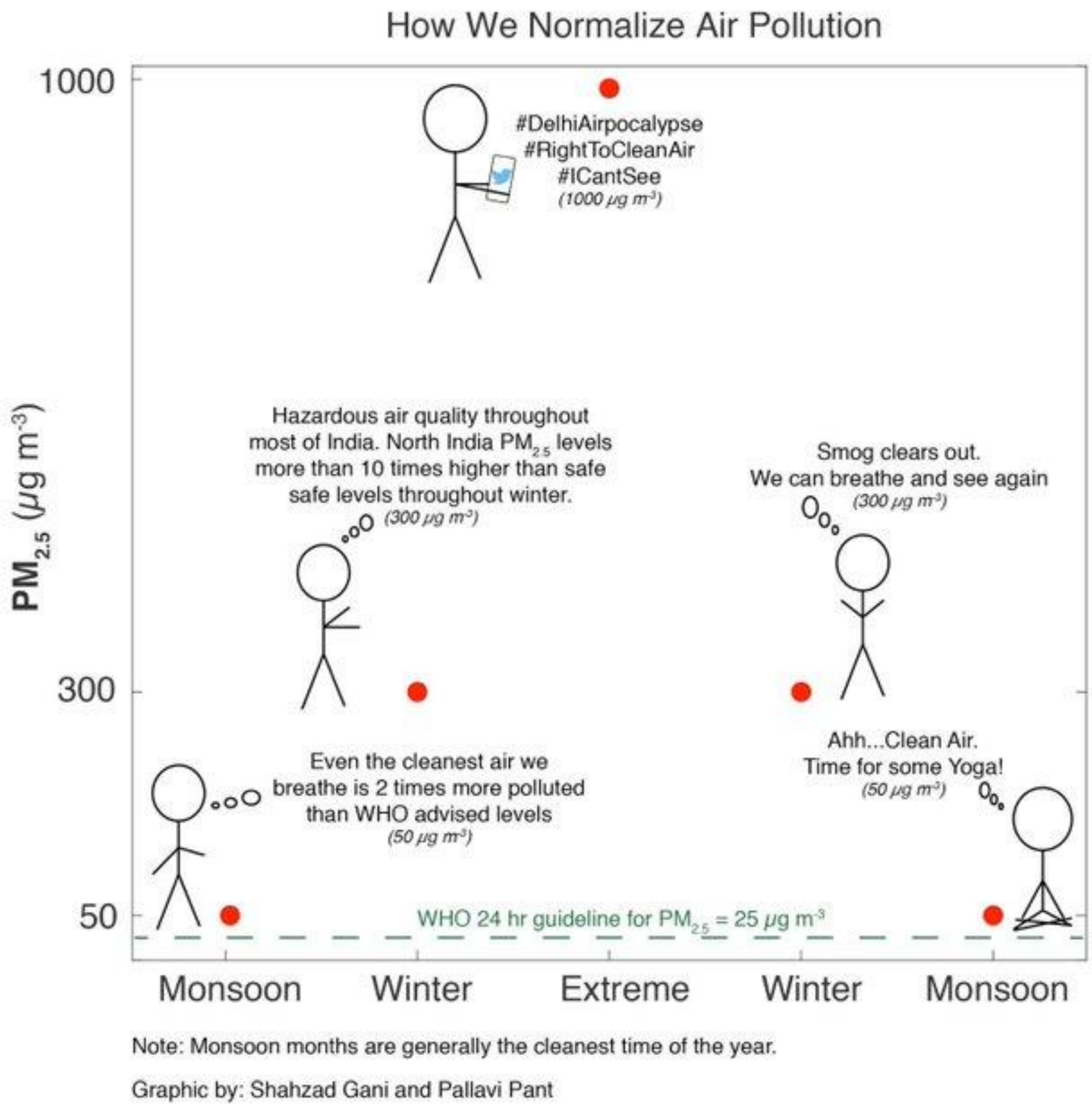


Figure: Percentage change in $PM_{2.5}$ exposure 2010-2019

Source: https://www.stateofglobalair.org/air/pm

Air pollution perspective



Source: Shahzad Gani and Pallavi Pant

Need of study

Concentrations on fixed sites are poorly correlated with personal exposure.

Ambient air pollution varies across the metropolitan environment, so to capture the spatial variation of air pollution, mobile monitoring is required.

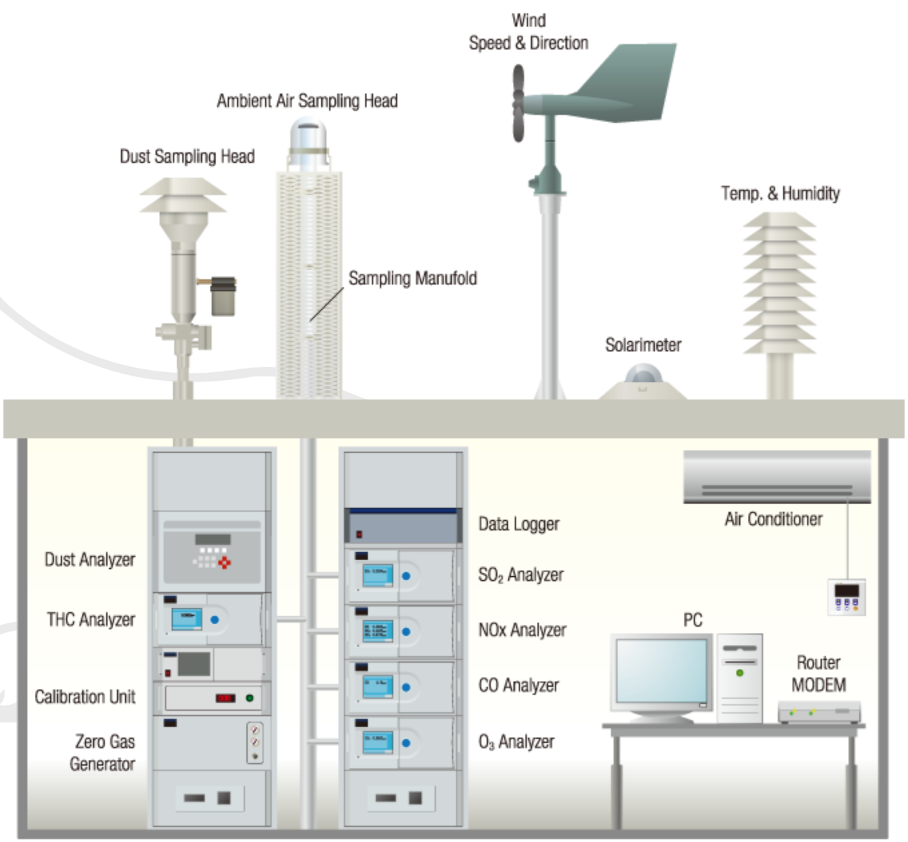
Travel is the one activity where people are exposed throughout the day.

The mobility is ignored that can bias the effect estimates of air pollution exposure

Air pollution monitoring

- Types of monitors:
  - Static monitors
  - Portable monitors
  - IoT based monitors
  - Mobile monitors

Portable monitor
(pocket size)



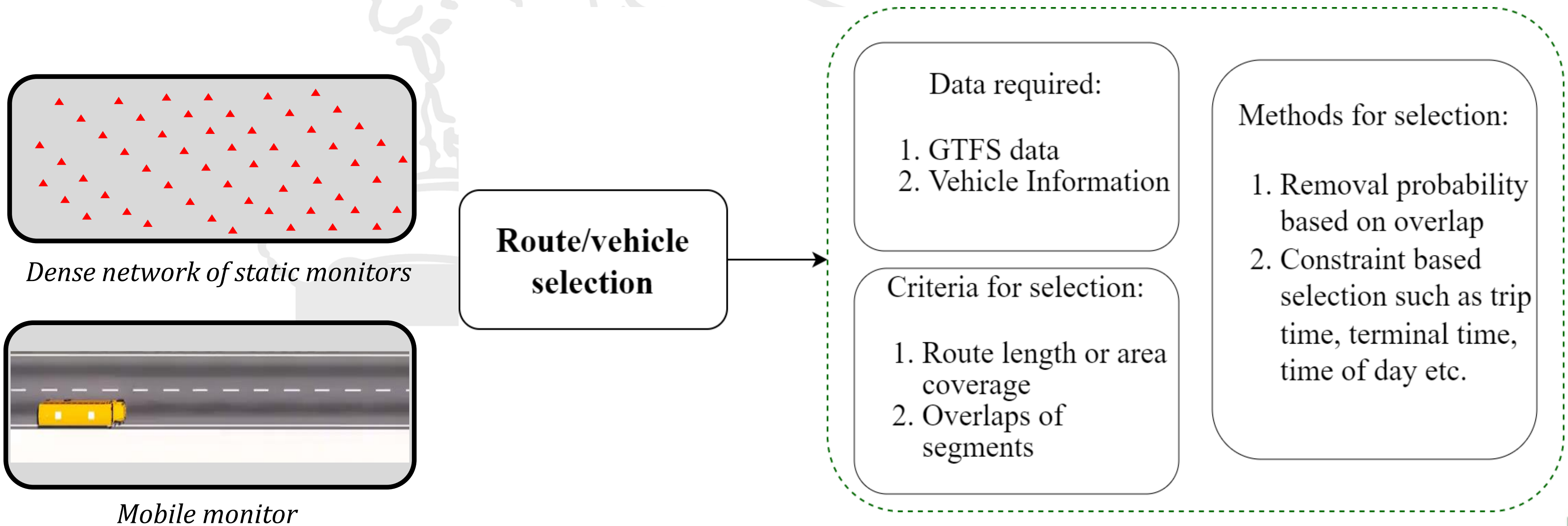
Static monitor

Location of mobile sensors

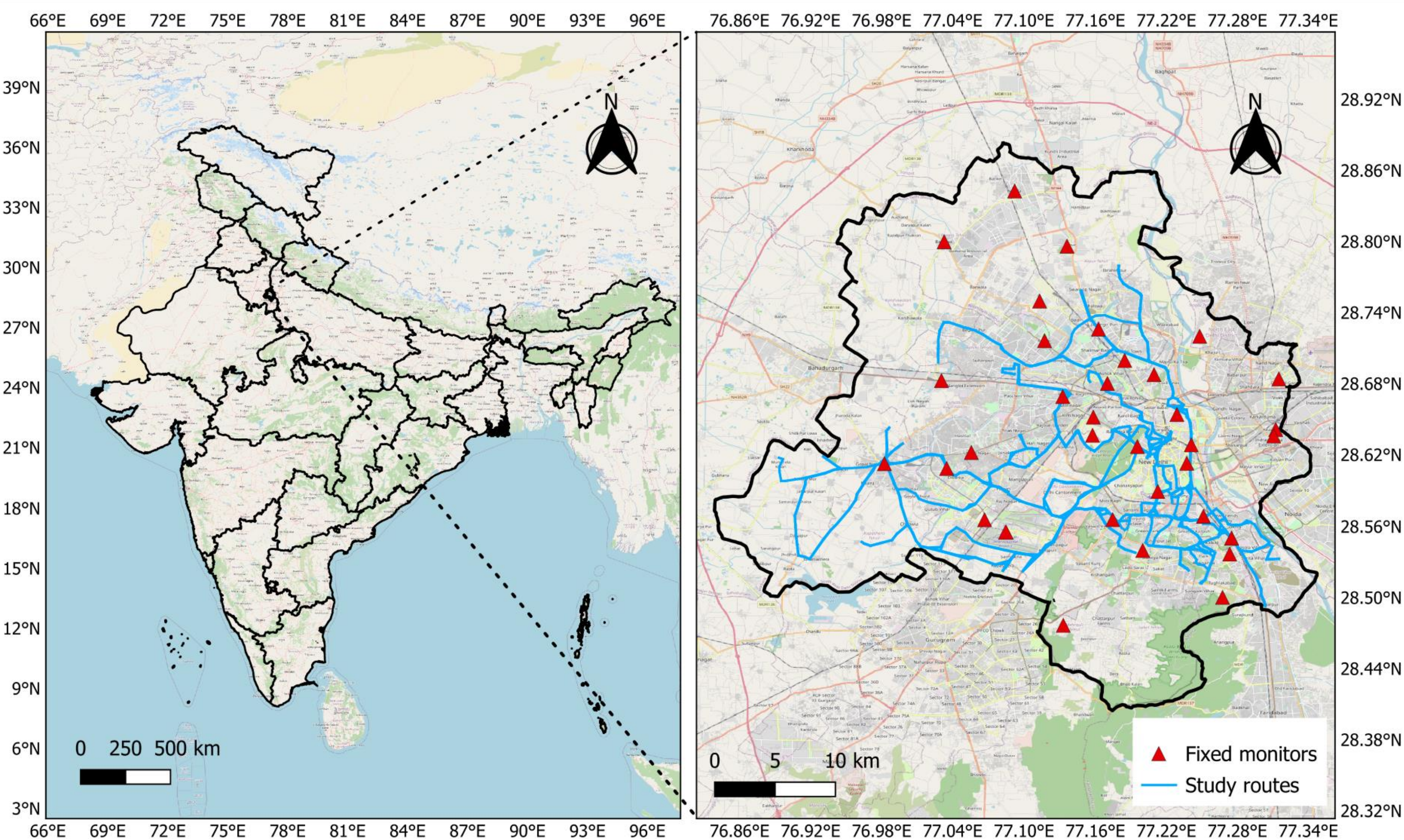
- Monitoring of air pollution is necessary to provide information about **daily exposure to researchers, policy makers, and city residents.**
- Dense **monitoring network of static monitors** is essential for accurate information.
- Portable monitors – higher number is required, which becomes **costly for larger cities**.
- Placement of monitors on **moving vehicles** modes can be a viable solution to collect real-time data covering larger area spatially.

Route selection for mobile monitoring

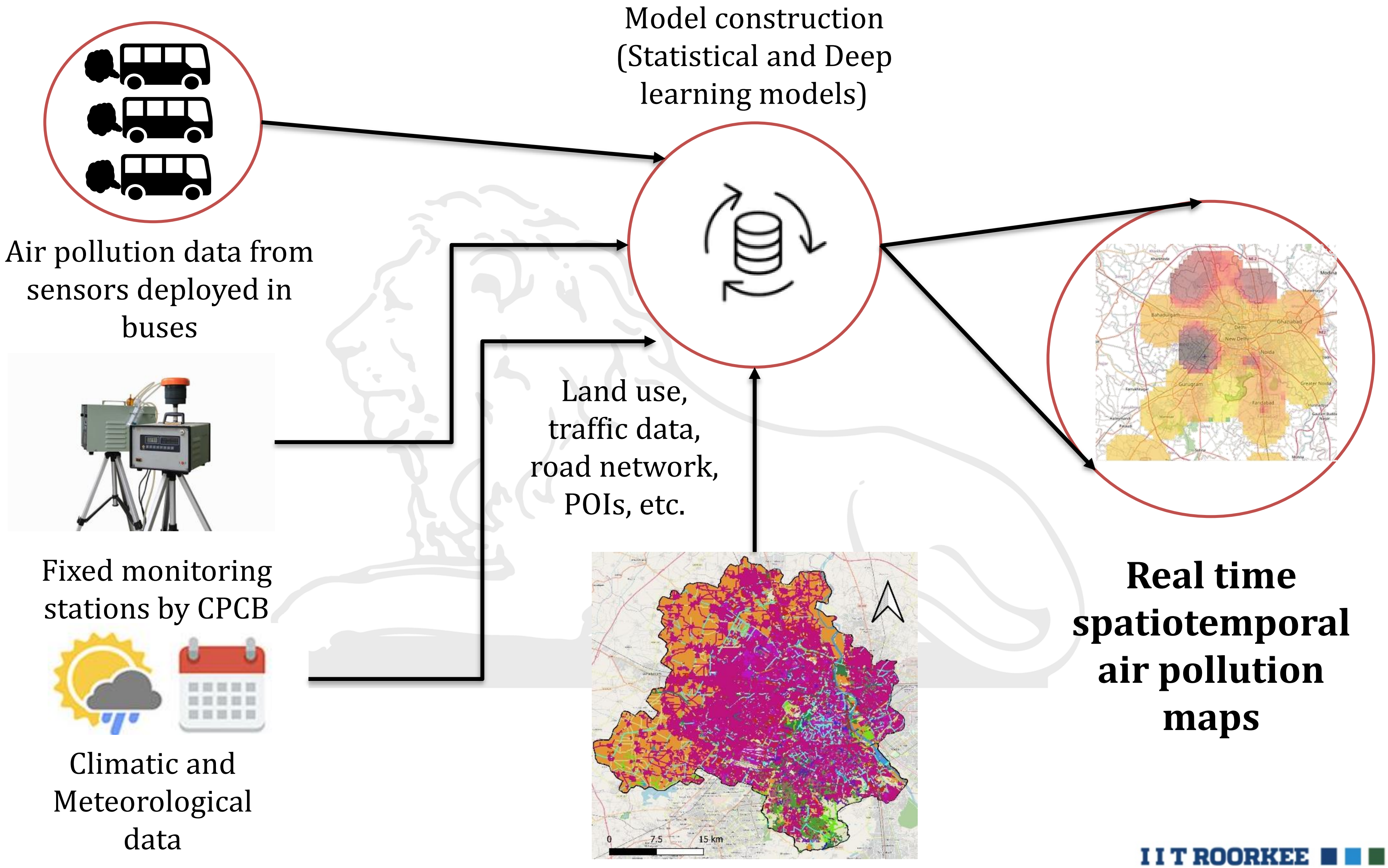
- Placement of monitors on **existing public transit** modes can be a viable solution to collect high resolution and real-time data.
- The constraints for selection process will be number of vehicles/portable monitors, **overlapping** of routes, **coverage area**, etc.
- A simulation algorithm is developed to select routes based on constraints.



Study area


66°E 69°E 72°E 75°E 78°E 81°E 84°E 87°E 90°E 93°E 96°E
39°N
36°N
33°N
30°N
27°N
24°N
21°N
18°N
15°N
12°N
9°N
6°N
3°N
N
0 250 500 km
76.86°E 76.92°E 76.98°E 77.04°E 77.10°E 77.16°E 77.22°E 77.28°E 77.34°E
28.92°N
28.86°N
28.80°N
28.74°N
28.68°N
28.62°N
28.56°N
28.50°N
28.44°N
28.38°N
28.32°N
N
0 5 10 km
Fixed monitors
Study routes

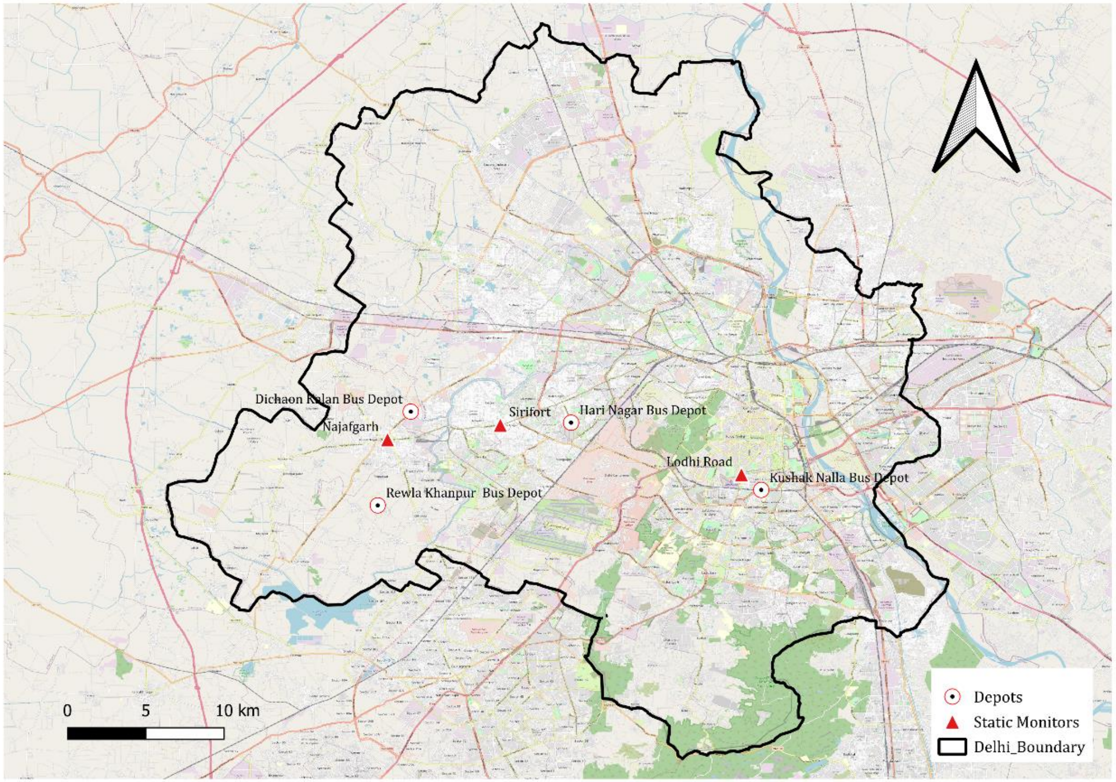
Mobile monitoring



Data collection details

- Date: August 21, 2022 to April 30, 2023
- Number of buses: 15
- Bus depots (Four):

Hari Nagar Depot (3 buses)
Kushak Nallah Depot (3 buses)
Dichaon Kalan Depot (4 buses)
Rewla Khanpur Depot (5 buses)



IIT ROORKEE

Data collection

Installation of monitors in buses

Data preprocessing

Following steps were used for data filtering:

- Remove duplicates values.
- Remove $PM_{2.5}$ values less than 5 µg/m^3 and more than 1000 µg/m^3.
- Use modified Z score to remove outliers.

Modified Z-score method to detect the outliers, which is a **more robust method** than all others used methods.

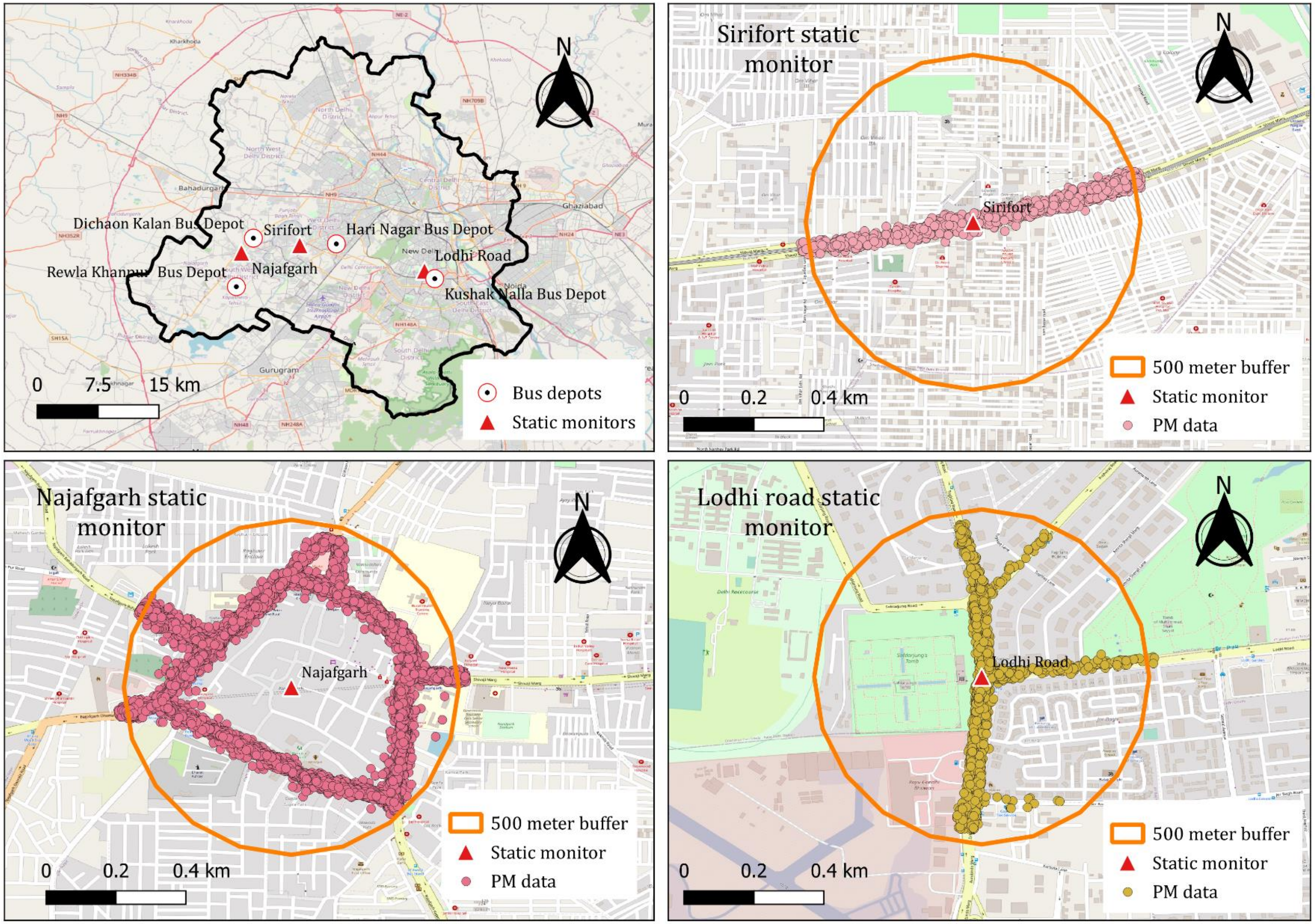
$$M_i = \frac{0.6475.(x - \bar{x})}{MAD}$$

where MAD is Median Absolute Deviation, $\bar{x}$ is the Median.

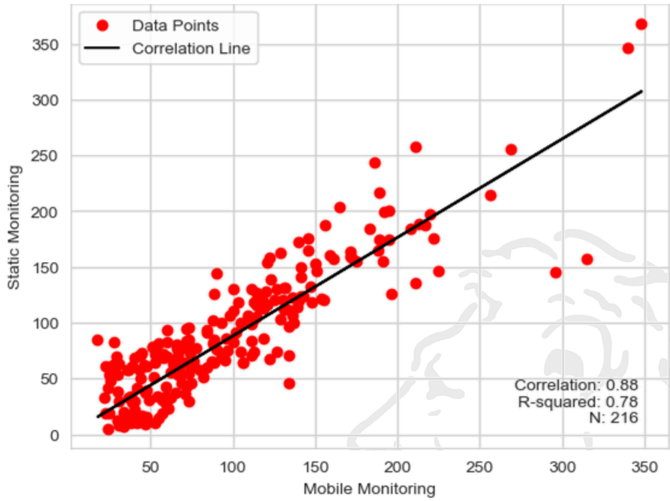
To remove the potential outliers from the dataset, $|M_i|$ **= 3.5** has been used (Iglewicz & Hoaglin, 1993).

Source: Singh and Agarwal, 2024.

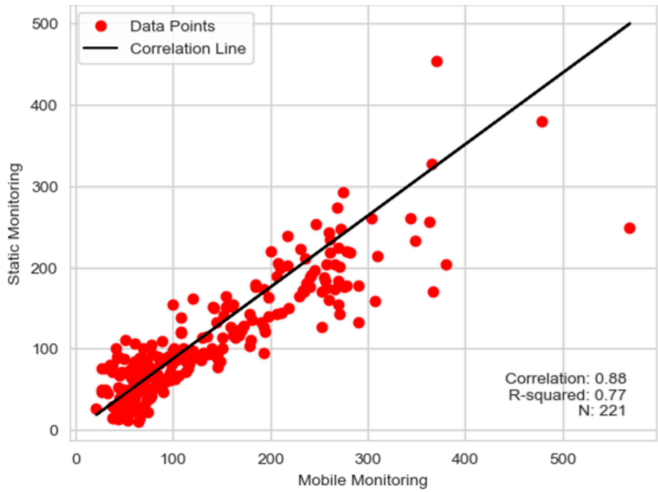
Fixed vs. Mobile data



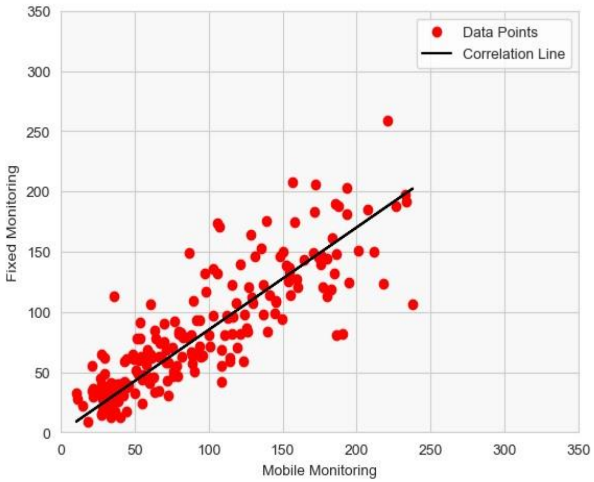
Fixed vs. Mobile



(a) Najafgarh

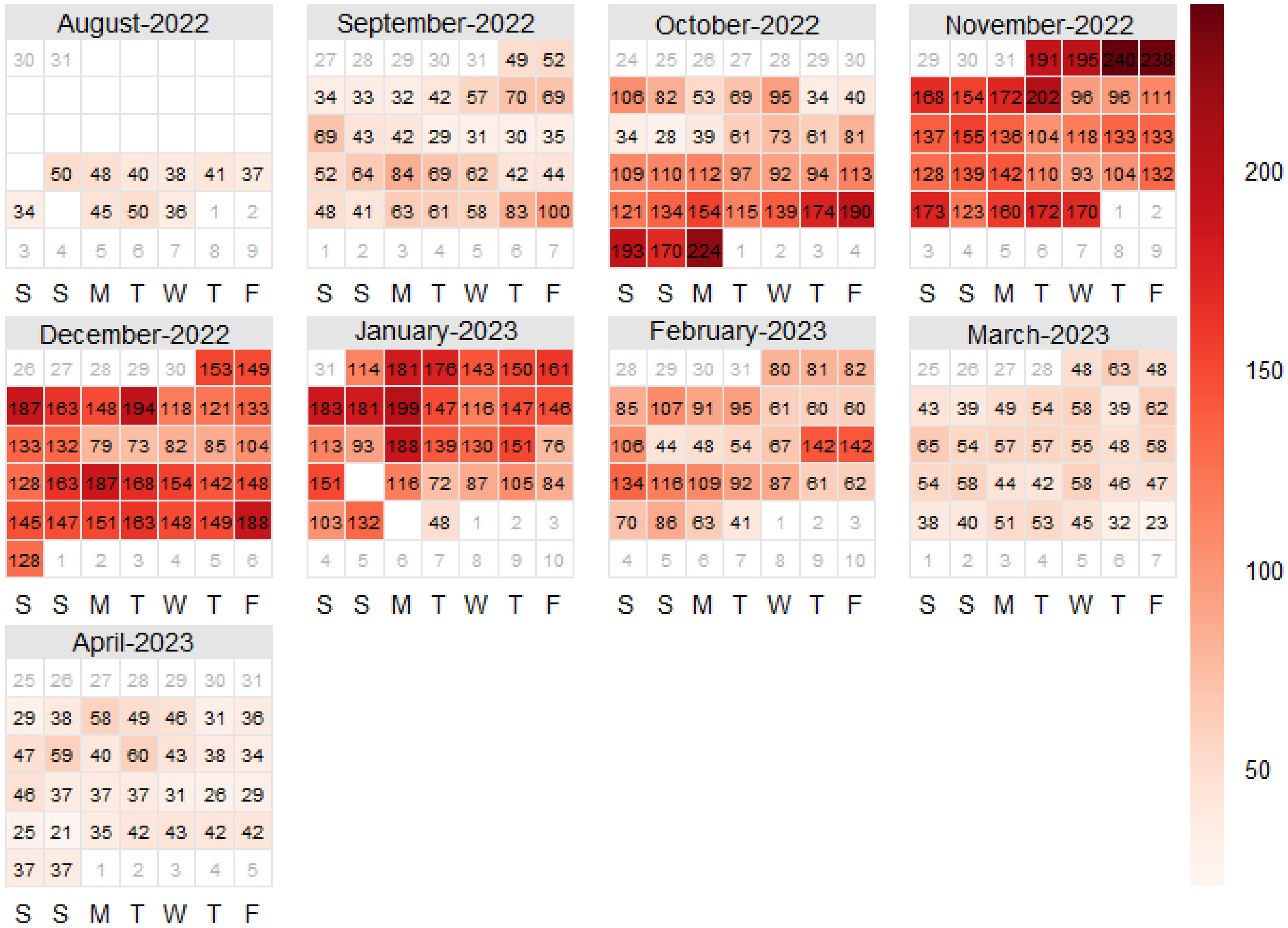


(b) Siri Fort

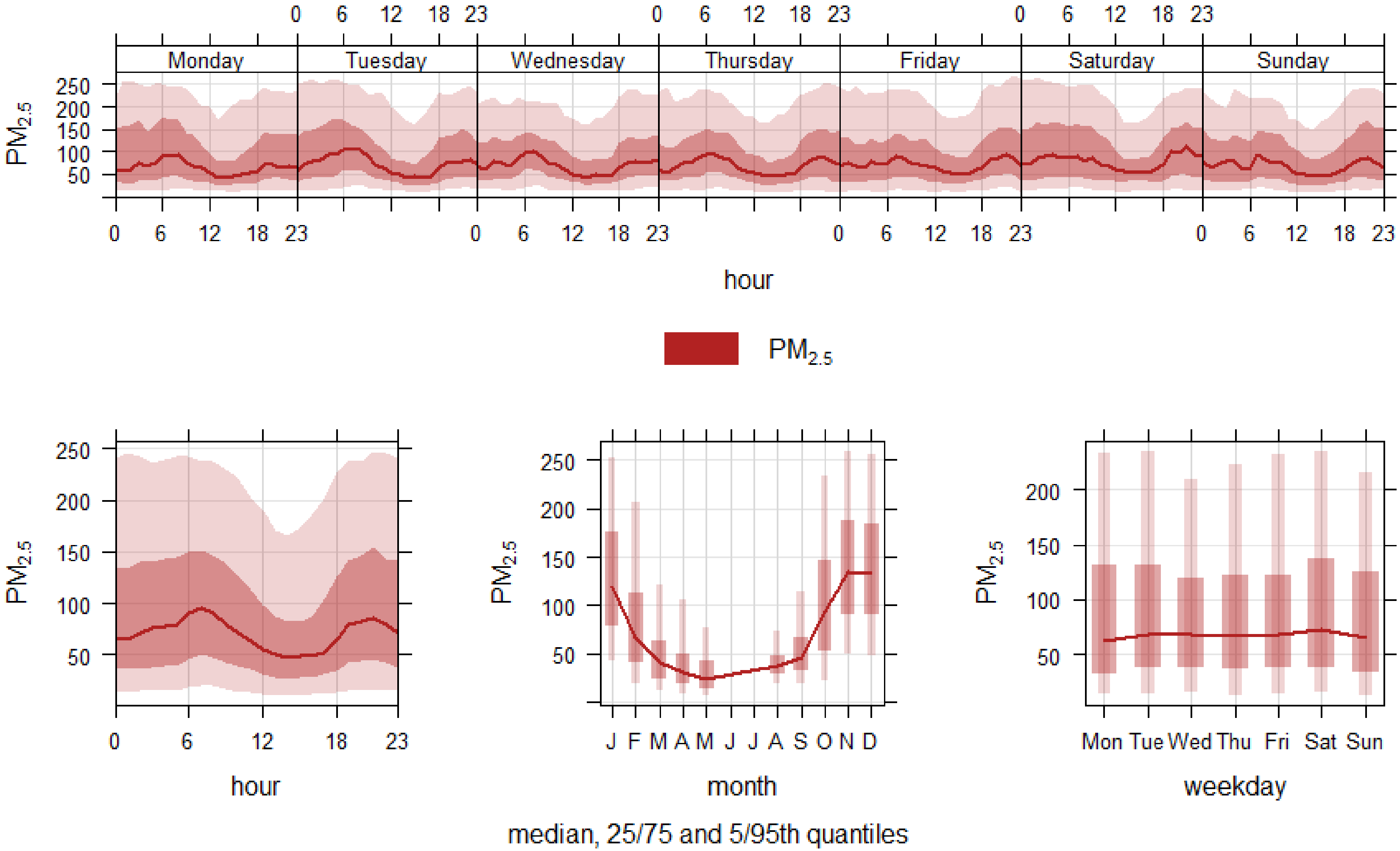


(c) Lodhi road

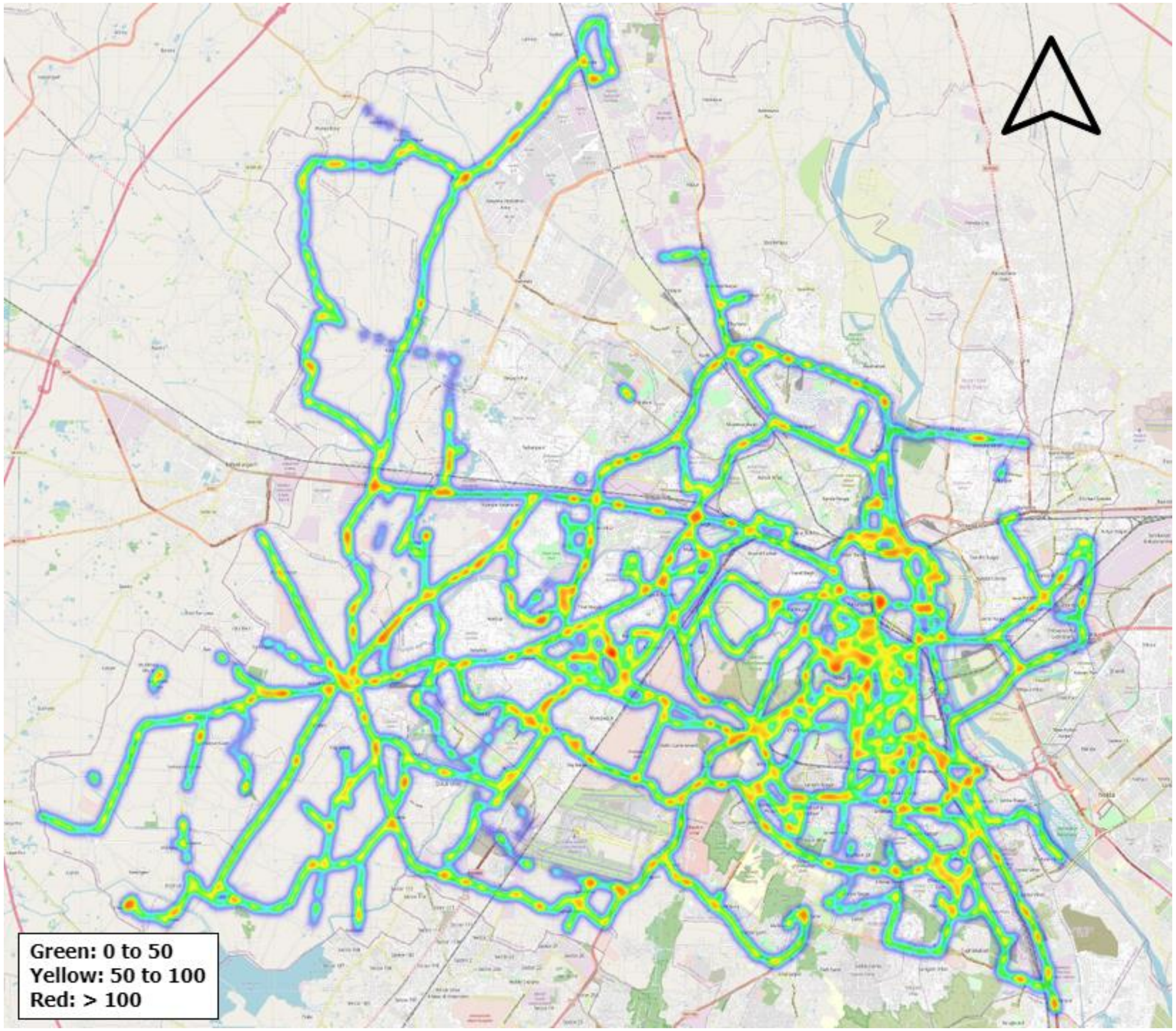
Temporal variations



Temporal variations



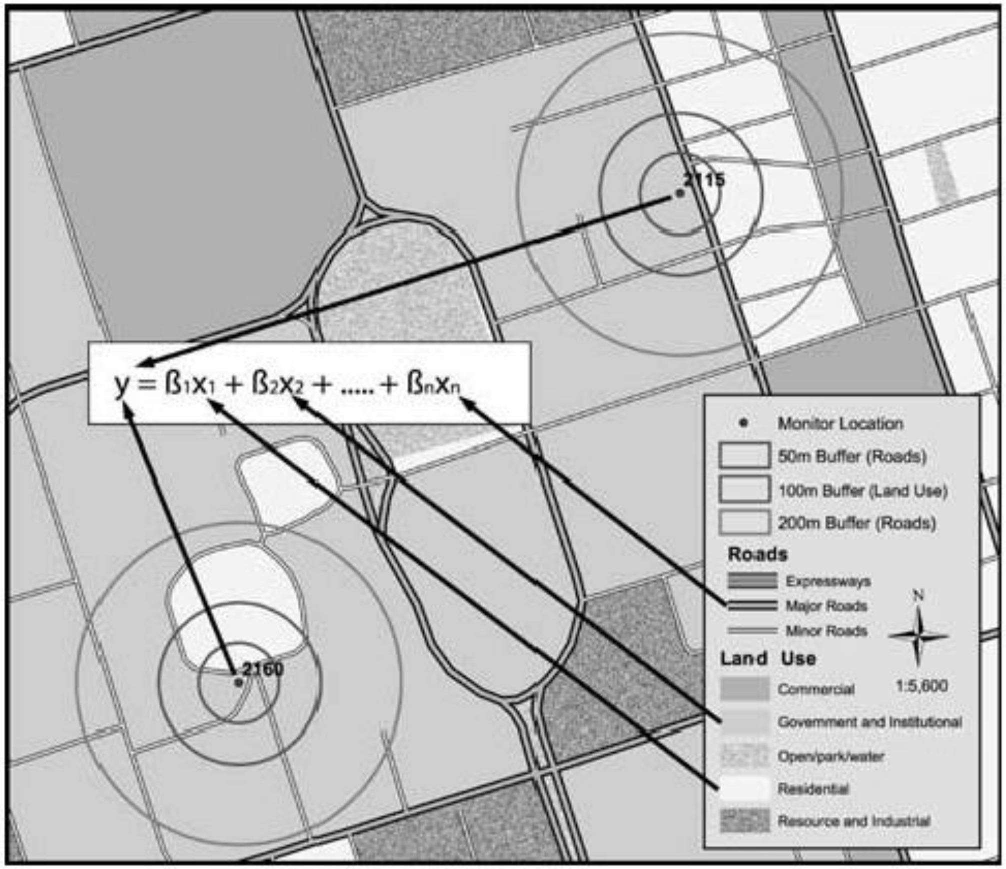
Spatial variations



Model formation

Each characteristic is assumed to be **linearly related** to **pollutant concentrations**.

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \cdots \ldots\ldots + a_nX_n$$



In which $X_1, X_2, X_3, \ldots, X_n$ are n samples of independent variables; Y is the dependent variable; a_0, a_1, a_2, a_3, ... a_n are coefficients of the independent variables.

Source: (Jerrett et. al, 2005)

Land cover

Categories	Abbreviations	Descriptions	Buffer	Unit	Prior direction
Land cover	Crop	Area of crops and grasses	TRUE	m^2	Neagtive
	Trees	Area of dense vegetation	TRUE	m^2	Neagtive
	Blt_Area	Area of buildings, impervious spaces, roads and rail networks	TRUE	m^2	Positive
	Br_Ground	Areas of rock or soil with very sparse to no vegetation	TRUE	m^2	Positive
	Water	Area of the water body	TRUE	m^2	Neagtive
	Fd_Veg	Areas of any type of vegetation with obvious intermixing of water	TRUE	m^2	Negative
	Rg_land	Area of open areas covered in homogenous grasses	TRUE	m^2	Negative

Source: https://livingatlas.arcgis.com/landcoverexplorer/

Land use

Categories	Abbreviations	Descriptions	Buffer	Unit	Prior direction
Land use	Residential	Area of residential land	TRUE	m^2	Positive
	Commercial	Area of commercial land	TRUE	m^2	Positive
	Industrial	Area of industrial land	TRUE	m^2	Positive
	Transport	Area of land for transportation facilities	TRUE	m^2	Positive
	Agriculture	Area for agriculture	TRUE	m^2	Negative
	Government	Area of government land	TRUE	m^2	Positive
	River	Area of river	TRUE	m^2	Negative
	PSP	Area of public and semipublic facility land	TRUE	m^2	Positive
	Recreational	Area of recreational land	TRUE	m^2	Negative
	Utility	Area of utilities	TRUE	m^2	Positive
	Spl_Area	Area of special area	TRUE	m^2	Positive

Source: https://download.geofabrik.de/asia/india.html

Road variables

Category	Abbreviation	Descriptions	Buffer	Unit	Prior direction
Roads and Traffic	Motorway	Length of the motorway roads	TRUE	m	Positive
	Primary	Length of the primary roads	TRUE	m	Positive
	Secondary	Length of the secondary roads	TRUE	m	Positive
	Tertiary	Length of the tertiary roads	TRUE	m	Positive
	Residentialroads	Length of the residential roads	TRUE	m	Positive
	Smallroads	Length of the small roads	TRUE	m	Positive
	Unclassified	Length of unclassified roads	TRUE	m	Positive
	Nonmotor	Length of the nonmotor roads	TRUE	m	Negative
	Dist2NrMotorway	Distance to the nearest motorway	FALSE	m	Negative
	Dist2NrPrimary	Distance to the nearest primary	FALSE	m	Negative
	BusStopNums	Number of bus stops	TRUE	m	Positive
	TrafficSignals	Number of traffic signals	TRUE	m	Positive

Source: https://download.geofabrik.de/asia/india.html

POIs

Categories	Abbreviations	Descriptions	Buffer	Unit	Prior direction
POIs	RestNums	Number of restaurants	TRUE	count	Positive
	EduNums	Number of educational institutions	TRUE	count	Positive
	MarkAreaNums	Number of market area	TRUE	count	Positive
	TouAttNums	Number of tourist attractions	TRUE	count	Positive
	DIS2NrRest	Distance to the nearest restaurants	FALSE	m	Negative
	DIS2NrMarkArea	Distance to the nearest market area	FALSE	m	Negative
	DIS2NrTouAtt	Distance to the nearest Tourist attractions	FALSE	m	Negative
	DIS2NrEdu	Distance to the nearest educational institution	FALSE	m	Negative

Source: https://download.geofabrik.de/asia/india.html

Building variables

Category	Abbreviation	Descriptions	Buffer	Unit	Prior direction
Buildings	BuildArea	Sum of area of building footprints	TRUE	m^2	Positive
	BuildVolume	Sum of built-up volume of the buildings	TRUE	m^3	Positive
	BuildHeight	Average height of the buildings	TRUE	m	Positive

Source (Height and Volume): https://ghsl.jrc.ec.europa.eu/datasets.php
Source (Area): https://sites.research.google/open-buildings

Other variables

Category	Abbreviation	Descriptions	Buffer	Unit	Prior direction
Population	Pop_Count	Population count	TRUE	Count/100 m^2	Positive
Meteorology	Temp	Temperature	FALSE	°C	Negative
Meteorology	RH	Relative humidity	FALSE	%	Positive
Elevation	DEM	Elevation at the midpoint of the segment	FALSE	m	Negative
Night Time Light	MeanNTL	Mean night time light value	TRUE	N/A	Positive

Source (Population): https://ghsl.jrc.ec.europa.eu/datasets.php
Source (Elevation): https://srtm.csi.cgiar.org/srtmdata
Source (Night Time Light): https://ladsweb.modaps.eosdis.nasa.gov

Mobile data preparation for modeling

Filter the relevant data from database(csv file).

↓

Then make buffer of 50 meter on roads meter file.

↓

Then make segments of 100, 500 meter or OSM roads using segments code.

↓

Then count the data points within the buffer of a segment.

↓

Then remove the segments with less number of data points (30).

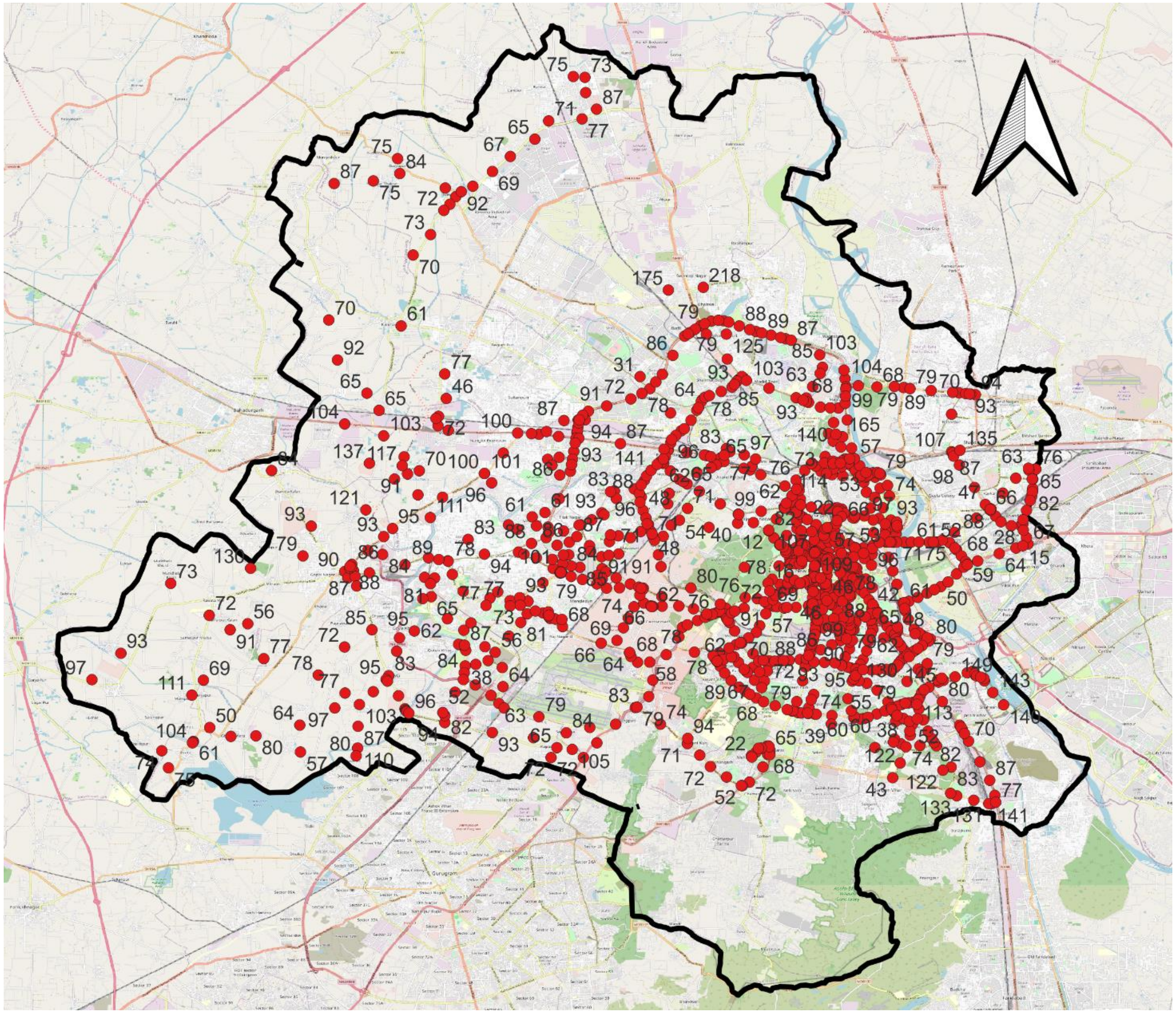
↓

Then snap the data points using snap option in QGIS.

↓

Then using segments and data point file calculate midpoint on segment.

Mid point of segment (Aggregation)



Linear regression models

Forward Linear Regression (FLR)

- FLR begins by building a model with no predictor variables (null model).
- At each step, it considers all remaining predictor variables and selects the one that, when included, leads to the highest increase in the model's adjusted R^2.
- Approach continues until no variable can considerably improve the modified R^2 value.

Backward Linear Regression (FLR)

- BLR starts with a model containing all predictor variables.
- It then iteratively removes the variable with the weakest association with the target variable (highest p-value) at each step.
- Process continues until remaining variables in the model are statistically significant (p-value < 0.1) and the model achieves the highest adjusted R^2 value.

Stepwise Linear Regression (WLR)

- It starts by building a model with no variables (null model) and iteratively adds the most significant predictor.
- It removes previously added variables if its p-value becomes greater than a threshold ($p = 0.1$).
- This back-and-forth process continues maximising adjusted R^2 and maintaining statistically significant predictors.

LUR model

First, the initial model was determined by a univariate regression of all predictors, and variables with the highest adjusted R^2 and consistent with the predefined direction is start of the model. A variable is added when all the following conditions are met:

1. the **adjusted R^2** of the model increased by **more than 1%,**
2. the **direction of the effect** of the newly added variable was consistent with the **predefined direction,**
3. the direction of the **coefficients of the original variables** in the model did not change,
4. Iterate **steps 2–3** until the additional increase of **R2 is < 1%** if all the remaining potential predictor variables are tried to update a 'new model'.
5. The predictor variables in the 'current model' with **p-value >0.10** are excluded.
6. Remaining variables are collinearity is checked by the **variance inflation factor** (VIF, any predictor variable with a **VIF >3.0** is not acceptable).

Generalized Additive Model (GAM)

The GAM model is used to fit this nonlinear relationship of all variables using the spline smoothing function. Generalized cross-validation was used to select the degree of smoothing.

The GAM is defined by:

$$g(\mu) = \beta_0 + s_1(x_1) + s_2(x_2) + s_3(x_3) + \cdots + s_p(x_p) + \epsilon$$

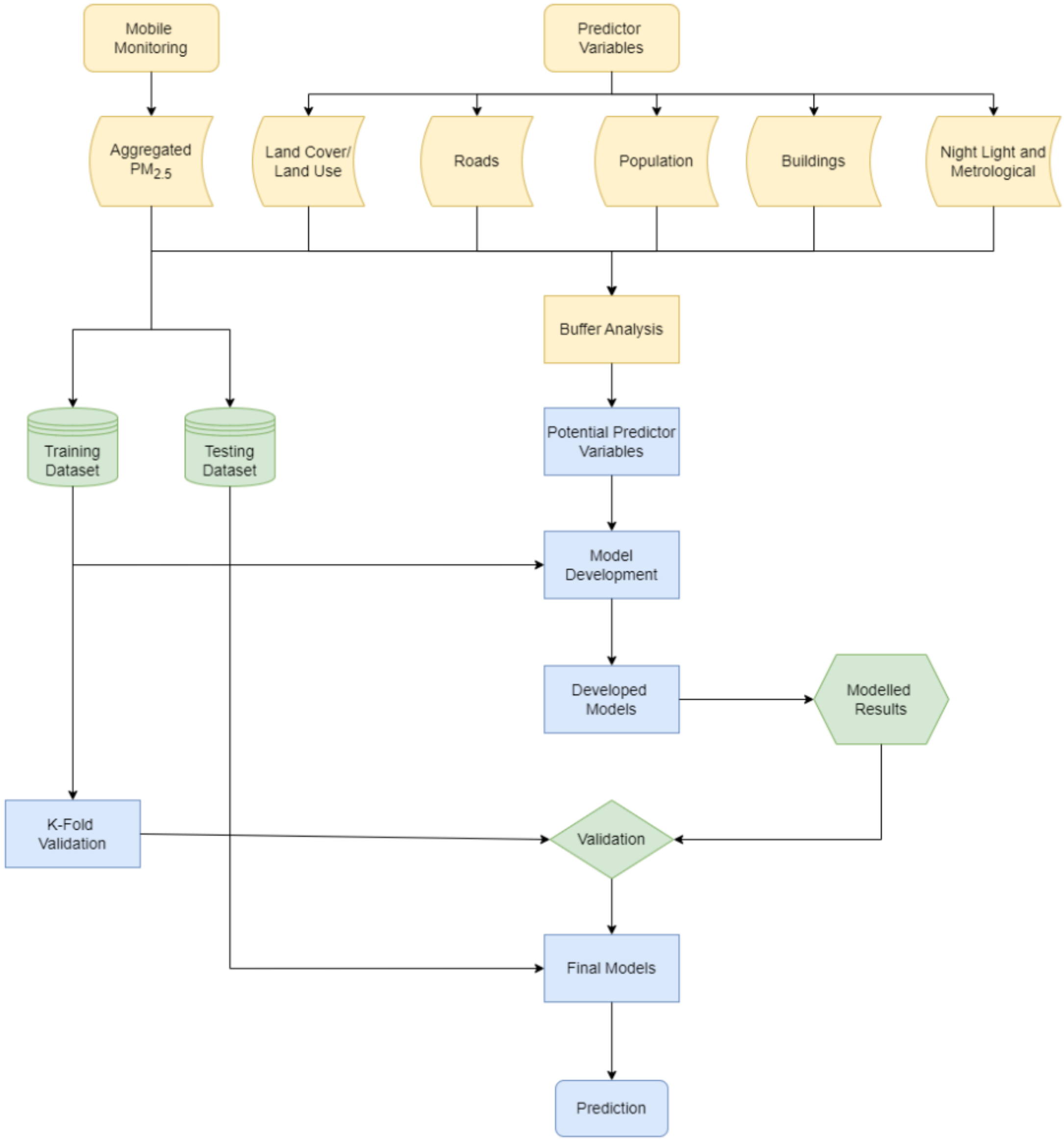
$$\mu = E\,(y|x_1, x_2, \cdots, x_p)$$

where g is the link function, and s_i is the spline smoothing function for each predictor variable.

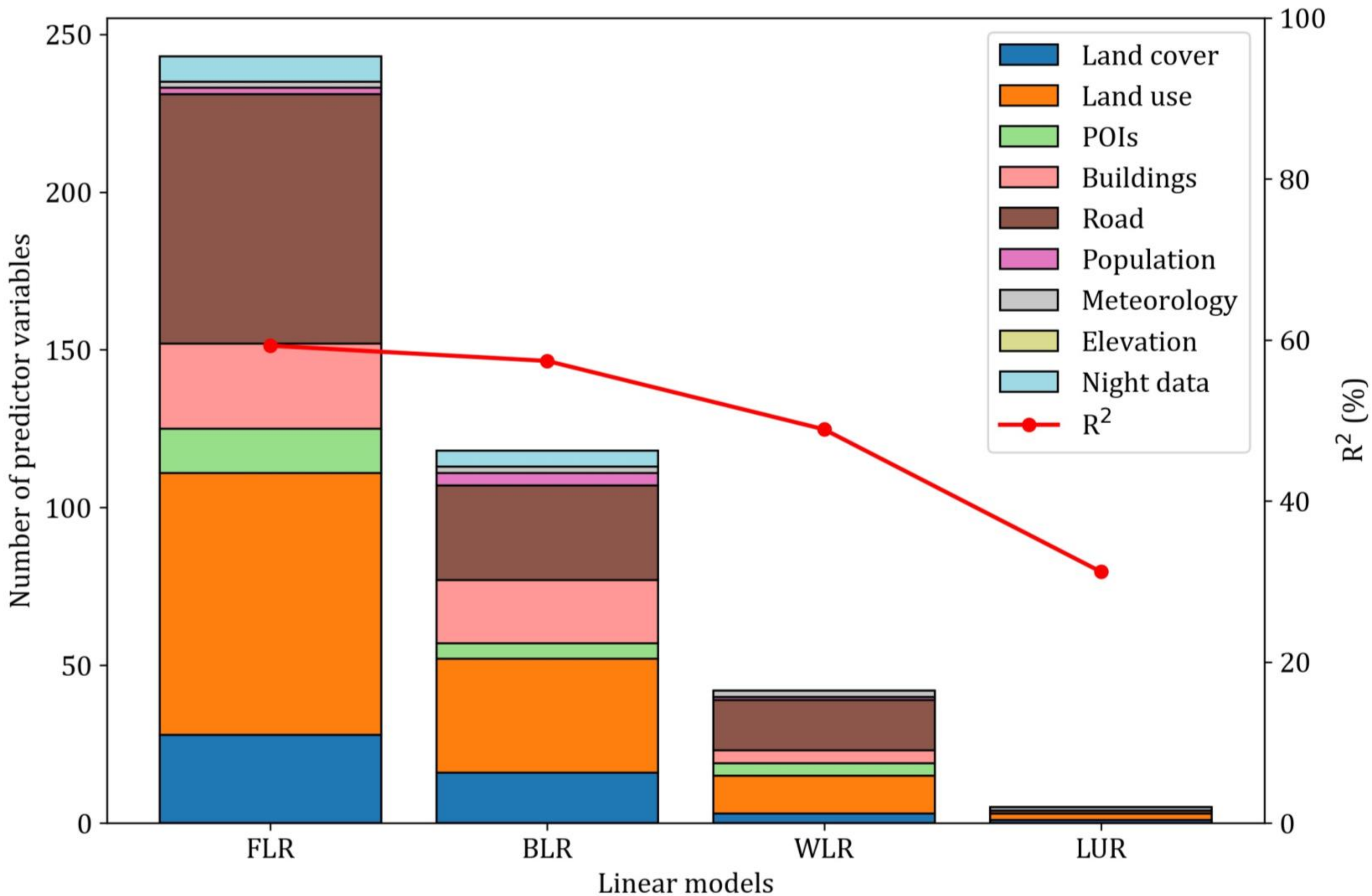
$s_p(x_p)$ is estimated for all s = 1,2......, p and then added together.

That's why this model is called an additive model.

Methodology



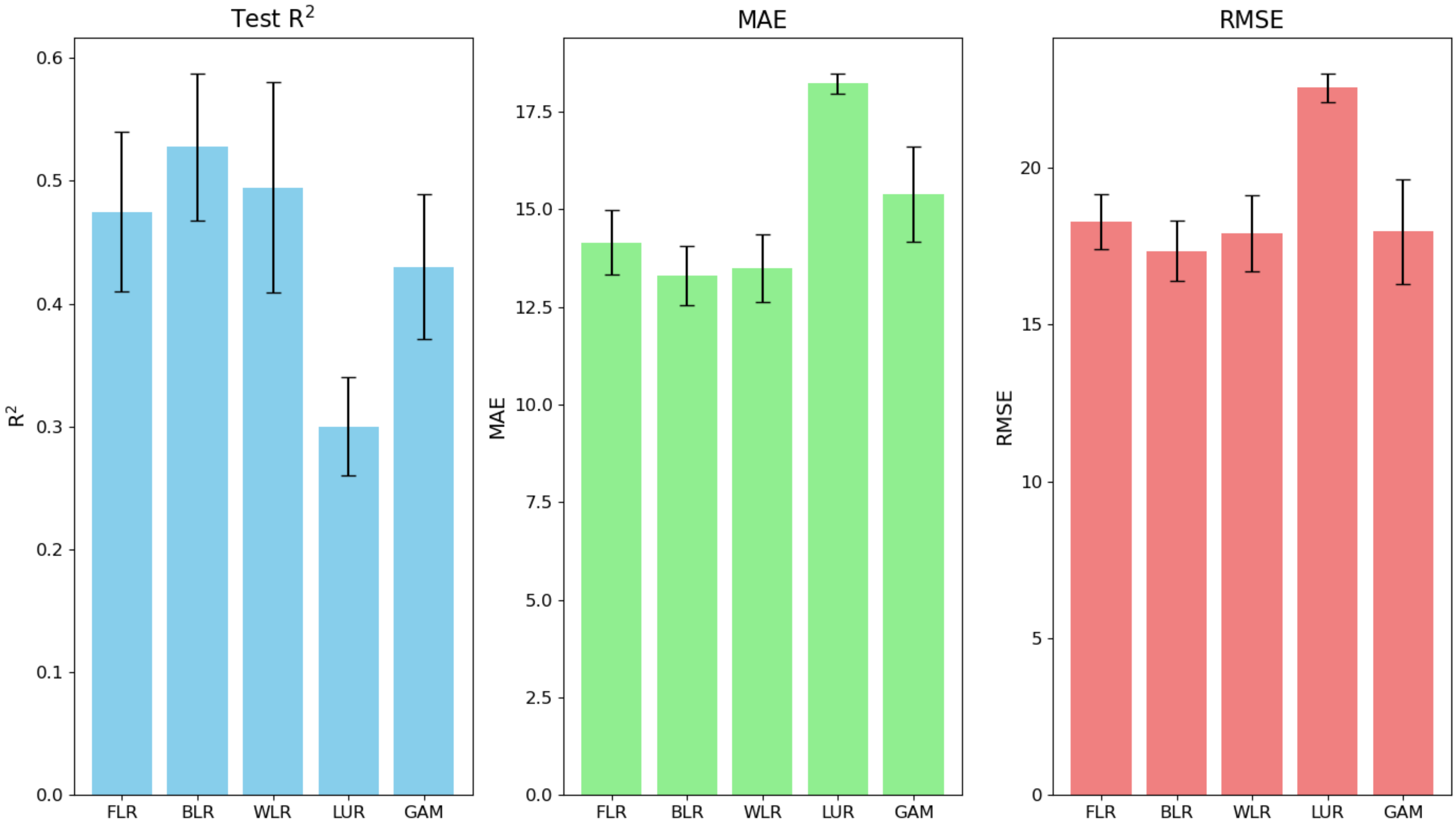
Variable retained



Model results

Type	Model	Training		
		R^2	MAE	RMSE
Linear	FLR	0.59	12.01	16.97
	BLR	0.57	12.41	16.58
	WLR	0.53	13.88	18.92
	LUR	0.31	16.05	18.55
Non linear	GAM	0.72	14.58	18.16

Model validation



Findings

- $PM_{2.5}$ concentrations have large **temporal and spatial** variations.
- **Temperature and relative humidity** are two most important parameters for model formation.
- Other major parameters are **building height, road network** characteristics (types of roads and traffic signals), **utility areas**, and distance to the **nearest highway**.
- **BLR** is best linear model as it removes insignificant variables and gives better performance and a **more accurate** model.
- Linear models has **higher prediction consistency** than non linear model.
- For **stable and consistent prediction** performance, linear models can be preferred and to **incorporate complex variables** in the model formation non linear models can be explored.

Conclusions

- Study uses **mobile monitoring** to measure $PM_{2.5}$ concentrations throughout Delhi.
- Data was collected using **15 low-cost air quality devices** for **eight months.**
- $PM_{2.5}$ concentrations have large **temporal and spatial variations**.
- Different models are used to model air pollution using variables such as **land cover/ land use, buildings, roads, and geographic and meteorology** variables.
- The linear algorithms did not perform better in **training dataset** because of a **nonlinear association** and the presence of **higher dimensional data**.
- The superior performance of nonlinear model in training dataset might be attributed to their capacity to **handle complex associations** between the data.
- These techniques offer significant advantages for studying **spatiotemporal fluctuations in air pollution.**

References

- K. R. Smith. , 1993. Fuel combustion, air pollution exposure, and health: The situation in developing countries. Annual Review of Energy and Environment, 18:529{566}.
- WHO, 1999. Monitoring ambient air quality for health impact assessment, WHO Regional Publications, European Series. World Health Organization - Regional Office for Europe, Copenhagen.
- Duan, N., 1991. Stochastic microenvironment models for air pollution exposure. Journal of Exposure Analysis and Environmental Epidemiology, 235-257.
- Klepeis, N.E., 2006. Modeling Human Exposure to Air Pollution, Human Exposure Analysis. CRC Press, Stanford, CA, pp. 1-18.
- Nethery, E., Teschke, K., Brauer, M., 2008. Predicting personal exposure of pregnant women to traffic-related air pollutants. Science of The Total Environment 395, 11-22.
- Piechocki-Minguy, A., Plaisance, H., Schadkowski, C., Sagnier, I., Saison, J.Y., Galloo, J.C., Guillermo, R., 2006. A case study of personal exposure to nitrogen dioxide using a new high sensitive diffusive sampler. Science of The Total Environment 366, 55-64.

References

- Kaur, S., Nieuwenhuijsen, M.J., Colvile, R.N., 2007. Fine particulate matter and carbon monoxide exposure concentrations in urban street transport microenvironments. Atmospheric Environment 41, 4781-4810.
- Briggs, D.J., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., Smallbone, K., 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. The Science of The Total Environment 253, 151-167.
- Boudet, C., Zmirou, D., Vestri, V., 2001. Can one use ambient air concentration data to estimate personal and population exposures to particles? An approach within the European EXPOLIS study. The Science of The Total Environment 267, 141-150.
- Choudhary, R., & Agarwal, A. (2024). Route selection for real-time air quality monitoring to maximize spatiotemporal coverage. Journal of Transport Geography, 115 , 103812.

References

- Brown, K.W., Sarnat, J.A., Suh, H.H., Coull, B.A., Spengler, J.D., Koutrakis, P., 2008. Ambient site, home outdoor and home indoor particulate concentrations as proxies of personal exposures. Journal of Environmental Monitoring 10, 1041-1051.
- K. J. Maji, A. Namdeo, D. Hoban, M. Bell, P. Goodman, S. M. S. Nagendra, J. Barnes, L. D. Vito, E. Hayes, J. Longhurst, R. Kumar, N. Sharma, S. K. Kuppili, and D. Alshetty. Analysis of various transport modes to evaluate personal exposure to PM2.5 pollution in Delhi. Atmospheric Pollution Research, 2020.
- Shiva Nagendra SM and Pavan Reddy Yasa and Narayana MV and Seema Khadirnaikar and Pooja Rani, 2019. Mobile monitoring of air pollution using low cost sensors to visualize spatiotemporal variation of pollutants at urban hotspots. Sustainable Cities and Society. 44, 520—535.
- Singh, V., & Agarwal, A. (2024). Variation of $PM_{2.5}$ and inhalation dose across transport microenvironments in Delhi. Transportation Research Part D: Transport and Environment , 127 , 104061.