CAPS
Centre for Air Pollution Studies

CSTEP®
CENTER FOR STUDY OF SCIENCE, TECHNOLOGY & POLICY

Mapping air pollution in Bengaluru using Low-Cost Sensors and Mobile Monitoring Data

Supported by

Google

Mapping air pollution in Bengaluru using low-cost sensors and mobile monitoring data

Center for Study of Science, Technology and Policy (CSTEP)
March 2023

The Center for Study of Science, Technology and Policy (CSTEP) is a private, not-for-profit (Section 25) Research Organisation registered in 2005.

ILK Labs (registered as ILK Labs LLP) is a research and educational organisation established in May 2017 with the goal to design and implement educational and research programmes on environment and sustainability.

Designed and edited by CSTEP

Disclaimer

Contributors: Dr Sreekanth Vakacherla (CSTEP), Dr Padmavati Kulkarni (CSTEP), Adithi R Upadhya (ILK Labs), Meenakshi Kushwaha (ILK Labs), Pratyush Agrawal (CSTEP), and Vinod Solomon (ILK Labs)

(*The author list provided assumes no particular order as every individual contributed to the successful execution of the project.*)

This report should be cited as follows: CSTEP & ILK Labs (2023). *Mapping air pollution in Bengaluru using low-cost sensors and mobile monitoring data* (CSTEP-RR-2023-3).

March 2023

Center for Study of Science, Technology and Policy

Bengaluru Office

18-19, 10th Cross, Mayura Street
Papanna Layout, Nagashettyhalli, RMV II Stage
Bengaluru-560094, Karnataka (India)
Tel.: +91 (80) 6690-2500

Noida Office

1st Floor, Tower A
Smartworks Corporate Park
Sector 125, Noida-201 303
Uttar Pradesh (India)

ILK Labs

62/H, 1st floor, Modi Residency
Millers Road, Benson Town
Bengaluru-560046, Karnataka (India)
Tel.: +91 (80) 4371-7396

Email: cpe@cstep.in
Website: www.cstep.in

Acknowledgements

We would like to take this opportunity to thank our funders and colleagues for their support and guidance throughout the monitoring campaign.

We thank Google for providing financial support to CSTEP for conducting low-cost sensor and mobile monitoring of air pollution studies in Bengaluru.

We thank Devaja Shah, Program Manager, Google, for her program management and technical guidance throughout the project tenure.

We also take this opportunity to acknowledge the CSTEP project leadership—Dr Jai Asundi, Dr Pratima Singh, and Dr Indu K Murthy—for their valuable support.

We thank Ajay Bhargav R (ILK Labs) and Mahesh G Kalshetty (CSTEP) for their help in the visualisation of results.

We thank the technical advisory members of the project Prof Julian D. Marshall (University of Washington), Prof John Volckens (Colorado State University), Prof Joshua S Apte (University of California, Berkeley), Prof Sagnik Dey (Indian Institute of Technology Delhi), Prof Cathryn Tonne (Barcelona Institute for Global Health), and Dr Nicholas Clinton (Google) for their valuable suggestions and technical guidance throughout the project tenure.

We thank the Central Pollution Control Board and the State Pollution Control Boards for installing and maintaining continuous ambient air quality monitoring stations and making the data publicly available.

We thank the Communications and Policy Engagement team of CSTEP for their editorial guidance and visual support.

We thank all the institutes and individuals who volunteered to host low-cost sensor on their premises.

Executive Summary

Air pollution is known for its detrimental health effects. Particulate matter (PM) pollution is one of the prominent drivers of the global burden of diseases (GBD) attributable to air pollution. To effectively manage air pollution, we need to measure it accurately and at high spatial resolution. However, maintaining a dense network of regulatory instruments is financially and technically burdensome for low- and middle-income countries. A hybrid approach that combines non-conventional, less expensive, short-term stationary, and mobile deployments may be a cost-effective solution. In the city of Bengaluru, India, we adopted such a hybrid measurement approach to generate high spatial resolution air pollution maps. We carried out a mobile monitoring campaign covering approximately 10% of roads in the city to measure on-road mass concentrations of fine particulate matter ($PM_{2.5}$), black carbon (BC), and number concentrations of ultrafine particles (UFPs). We also conducted another campaign where we established and maintained a 55-node city-wide network of low-cost sensors to measure ambient $PM_{2.5}$. Data from these two campaigns were corrected for their respective instrument biases and then used along with regulatory data from pollution control board monitors to predict pollutant levels at 50 m resolution using land-use regression models.

Based on the mobile measurements, we found that on-road air pollution levels (for all three pollutants) were highest on major roads, followed by arterial and residential roads. The spatial gradient in $PM_{2.5}$ (across various road types) was shallower compared to that of BC and UFPs. Also, the spatial representativeness of pollution maps increased when the number of drives on a route increased. On-road pollutant levels were higher than ambient pollution levels by a factor of 2 to 10. Ambient $PM_{2.5}$ levels (and their spatial variability) were higher in winter compared to summer. Daily mean ambient $PM_{2.5}$ levels were within Indian regulatory standards. We also evaluated the performance of various low-cost PM sensors in measuring $PM_{2.5}$.

Several important variables emerged from the land-use regression (LUR) models as potential predictors of the spatial variation of ambient $PM_{2.5}$. For example, area and line sources emerged as potential predictors for daytime on-road air pollution.

The predicted spatial ambient mean $PM_{2.5}$ was comparable across the Bengaluru Rural district (ranging between 33 μg m^{-3} and 42 μg m^{-3}) and the Bengaluru Urban district (35 μg m^{-3} to 44 μg m^{-3}), with little rural–urban contrast. The daytime on-road predicted $PM_{2.5}$ exhibited higher levels (between 72 and 80 μg m^{-3}) in the western parts of the Bruhat Bengaluru Mahanagara Palike (BBMP) area (including Dasarahalli, Rajarajeshwari Nagar, West, South, and parts of Bommanahalli) and very high concentrations (>88 μg m^{-3}) on major roads (Bengaluru–Mysore Road, NICE Road, Kanakpura Road, Magadi Main Road, and Tumkur Road). The daytime on-road predicted BC and UFPs followed a similar pattern (lower concentrations along BBMP boundaries and higher concentrations in South and West zones and along major and arterial road networks).

This work developed high-resolution maps of predicted ambient air pollution levels. The maps reveal that peri-urban and rural districts of Bengaluru have almost similar pollution levels. Regional mitigation action plans are, therefore, needed to effectively abate ambient air pollution. The primary policy recommendations from the work are (1) regulatory monitors need to be installed in non-urban areas, (2) supplementary monitoring using LCSs needs to be pursued in select regions in addition to the regulatory monitoring in urban areas, and (3) a mitigation plan is needed to reduce traffic-related emissions in western Bengaluru.

Contents

Figures

Tables

Abbreviations

Abbreviation	Expansion
AOD	Aerosol Optical Depth
BC	Black Carbon
BAM	Beta Attenuation Monitor
BBMP	Bruhat Bengaluru Mahanagra Palike
CAAQMS	Continuous Ambient Air Quality Monitoring Station
CPCB	Central Pollution Control Board
CNG	Compressed Natural Gas
CPC	Condensation Particle Counter
CSTEP	Center for Study of Science Technology and Policy
CV	Cross-Validation
DT	DustTrak Monitor
DJF	December, January, and February
DJFMAM	December, January, February, March, April, and May
ESCAPE	European Study of Cohorts for Air Pollution Effects
ESA	European Space Agency
FEM	Federal Equivalent Method
FRM	Federal Reference Method
GPS	Global Positioning System
Hz	Hertz
IoT	Internet of Things
km	Kilometres
KSPCB	Karnataka State Pollution Control Board
LCSs	Low-Cost Sensors
LMIC	Low- and Middle-Income Countries

LOOCV	Leave-One-Out Cross-Validation
L/min	Litres Per Minute
LUR	Land-Use Regression
LULC	Land Use and Land Cover
mA	microAeth
µg	Microgram
NCAP	National Clean Air Programme
NRMSE	Normalised Root Mean Square Error
NDVI	Normalised Difference Vegetation Index
NO_2	Nitrogen dioxide
nm	Nanometre
OSM	OpenStreetMap
PA	PurpleAir Monitor
PCB	Pollution Control Board
PM	Particulate Matter
$PM_{2.5}$	Particulate Matter With a Diameter of 2.5 Micrometres or Less
PM_{10}	Particulate Matter With a Diameter of 10 Micrometres or Less
R^2	Coefficient of Determination
RH	Relative Humidity
RMSE	Root Mean Square Error
T	Temperature
UFPs	Ultrafine Particles
UPAS	Ultrasonic Personal Air Sampler
USEPA	United States Environmental Protection Agency

1. Introduction

Being home to some of the most polluted cities in the world, India bears a very high burden of diseases and deaths associated with air pollution exposure. In 2019 alone, 1.67 million deaths were attributed to air pollution, accounting for ~18% of deaths in the country (Pandey et al., 2021). The resulting economic loss due to lost productivity was around 1.35% of India's gross domestic product (GDP) for the year. While the last two decades have seen a decline in deaths due to household air pollution, those due to ambient (or outdoor) particulate air pollution have increased substantially. Depending on the region, dominant sources of air pollution vary between industrial, household, energy production, and transportation sectors.

In India, particulate matter (PM) is one of the major drivers of morbidity and mortality attributable to air pollution. Because of its small size, $PM_{2.5}$ (i.e., fine particles with aerodynamic diameter less than 2.5 micrometres) can enter and deposit deep in our lungs, eventually reaching our bloodstream. Chronic exposure to $PM_{2.5}$ contributes to the risk of cardiovascular and respiratory diseases. In 2019, long-term exposure to $PM_{2.5}$ caused nearly 1 million deaths in India (State of Global Air, 2020). While $PM_{2.5}$ mass concentration is a criteria pollutant, other PM constituents are also a growing health concern. Black carbon (BC; a component of PM), for instance, is produced from combustion—major sources being combustion engines (mainly diesel), coal-based power plants, biomass cooking, and agricultural waste burning. Many health outcomes linked to $PM_{2.5}$ are also associated with BC (Janssen et al., 2011). Recent evidence shows that ambient BC can enter the foetal side of the placenta, exposing the developing foetus to air pollution (Bove et al., 2019). Another important metric of PM air pollution is number concentrations of ultrafine particles (UFPs; i.e., particles with size less than 100 nanometres [nm]). UFPs are either emitted directly from a source or formed in the atmosphere by photochemical aging of precursor gases. These particles have not been studied as extensively for their health effects as $PM_{2.5}$, but evidence regarding their detrimental health effects is beginning to accumulate (Ohlwein et al., 2019). Given their size, UFPs, upon depositing in the respiratory track, can eventually reach many of the body's organs, such as the liver, kidney, brain, and gastrointestinal tract (Schraufnagel, 2020). Short-term exposure to UFPs is associated with increased blood pressure and pulmonary inflammation, thus elevating the risk of cardiovascular diseases.

Over the years, the Government of India has implemented several programmes to monitor and control air pollution. Under the National Ambient Air Quality Monitoring Programme (NAMP), India has ~800 monitoring locations covering 344 cities/towns. The National Clean Air Programme (NCAP) was unveiled in 2019 with the goal of improving air quality in 122 target cities. One of the key action points of NCAP is to increase ambient air quality monitoring (Ganguly et al., 2020). Currently, India's monitoring capacity falls well below that of similar highly populated countries (Brauer et al., 2019). India has less than two monitors per 10 million people as opposed to China (12 monitors per 10 million people), the USA (34 monitors per 10 million people), and Brazil (18 monitors per 10 million people).

Existing monitoring stations in India are concentrated in urban areas, leaving many parts of the country routinely under-monitored or without monitoring. This data gap and the high cost of purchasing and maintaining reference-grade monitors call for integrated approaches that combine conventional regulatory monitoring with innovative and affordable solutions, such as low-cost sensors (LCSs) and periodic mobile monitoring. Such a hybrid approach has the potential

to enhance the current monitoring capacity and bridge critical data gaps for effective air quality management[1].

Hybrid systems that use LCS networks in combination with mobile monitoring of on-road pollution concentrations are gradually being adopted by cities around the world to obtain a comprehensive picture of neighbourhood-level variations in air pollution. These hybrid systems have been shown to provide actionable data that can inform policy decisions. One such example is the Breathe London project, where mobile monitoring data are combined with stationary LCSs to generate high-resolution pollution maps, identify pollution hot spots in London, and assess the impact of pollution control measures[2]. Google Street View cars fitted with air pollution monitoring equipment have been used to capture high-resolution pollution data in several cities in Europe and US, such as Oakland, Copenhagen, Amsterdam, and Dublin.

Though these monitoring techniques are gaining popularity globally, they are still relatively new to low- and middle-income countries (LMICs). Because of the unique challenges posed by low-resource settings (e.g., high pollution concentrations and poor infrastructure), these innovations need to be evaluated for feasibility and scalability before they are integrated into air quality management programmes in LMICs. To achieve this goal, this project combined two innovative approaches—a dense network of stationary LCSs and mobile monitoring of on-road pollution—to produce high-resolution air pollution data for the city of Bengaluru. The project helps test the feasibility and scalability of such an integrated approach for air quality monitoring in a prominent Indian urban centre. While mobile monitoring and LCS studies have been conducted in Indian urban centres before (Shiva Nagendra et al., 2019; CSTEP & ILK Labs, 2020), this project is the first of its kind to combine both modalities in a single geographic area to develop high-resolution pollution maps and estimate city-wide pollution levels using land-use regression models[3].

This study built on an existing low-cost PM sensor network in the Bengaluru Metropolitan Area that was established in 2019 via a collaboration between the University of Texas, the Center for Study of Science, Technology and Policy (CSTEP), and ILK Labs. We expanded the prior network (40 nodes) to a 55-node city-wide network. The new network covers a majority of the city area, major land-use types, and some peri-urban and rural locations around Bengaluru. Given that the market is flooded with LCSs that measure PM with varying levels of performance, we evaluated some of the most popular LCSs and documented best practices for conducting an LCS network project. The project team carried out a mobile monitoring campaign in several neighbourhoods and covered multiple road types—major, arterial, and residential—optimising protocols over an extended period of time. For the project, we covered over 1000 km of Bengaluru roads (around 10% of the total length of city roads) and collected unprecedented levels of air quality data, giving unique insights into the city's air pollution both spatially and temporally.

[1] https://www.vitalstrategies.org/resources/accelerating-city-progress-on-clean-air-innovation-and-action-guide/

[2] https://www.c40knowledgehub.org/s/article/The-Breathe-London-Blueprint-Supporting-cities-air-pollution-monitoring-goals?language=en_US

[3] https://blog.google/products/maps/100-million-air-quality-measurements-with-air-view/

Extensive air pollution monitoring activities under this project started in December 2021 and continued until May 2022. In this report, we present results obtained from data collected during these six months.

For simplicity, the terms $PM_{2.5}$, BC, and UFPs are used to indicate mass concentration of particulate matter (μg m^{-3}) having size less than 2.5 microns, black carbon mass concentration (μg m^{-3}), and number concentration of ultrafine particles (cm^{-3}), respectively, in the rest of the report.


0.20
µSv/h

2. Objectives of the Study

In this study, we aim to develop high-resolution pollution concentration maps in Bengaluru Urban and Rural districts using on-road measurements and land-use regression (LUR) models. We then analysed the resultant pollution estimates to produce a set of air pollution management and policy recommendations. The specific objectives of the study are as follows:

1. Expand and maintain the existing PurpleAir low-cost PM sensor network in Bengaluru, India.
2. Carry out performance comparisons of various PM LCSs.
3. Develop air pollution models by utilising heterogeneous data sources:
   a. mobile monitoring and
   b. stationary low-cost network.
4. Generate high-resolution $PM_{2.5}$, BC, and UFPs maps for public awareness and actionable recommendations.

3. Instrumentation and Data Handling

We measured daytime on-road $PM_{2.5}$, BC, and UFPs as part of the mobile monitoring campaign, and network measurements of ambient $PM_{2.5}$ were carried out using LCSs. A suite of instruments was used to measure on-road and ambient pollution and meteorological parameters (see Figures 1, 2, and 3). Instruments measuring on-road pollutant concentrations were configured to collect data at 1 Hz logging and averaging frequency. All instruments were factory-calibrated prior to the initiation of the measurement campaign and regularly monitored for their performance and reliability. Reference data on ambient air pollution were taken from the regulatory grade instruments installed on the CSTEP roof terrace at a height of ~10 m above ground level and ~110 m away from the main road.

3.1 On-road Air Pollution Measuring Equipment and Mobile Platform

3.1.1 Garmin

A GPSMAP 64s (Garmin Ltd., USA) was used to record location information (latitude and longitude) of the mobile monitoring vehicle. GPSMAP 64s is a high-sensitivity Global Positioning System (GPS) receiver with a quad helix antenna. The instrument works on the principle of trilateration to provide accurate location information. More details on this device can be found at garmin.com.

3.1.2 DustTrak

A DustTrak (DT) DRX Aerosol Monitor (model: 8533, TSI Incorporated, Minnesota, USA) was used to measure on-road $PM_{2.5}$. It is a portable, battery-powered instrument that can measure both mass and size fraction. It detects PM based on the optical scattering technique, operates at a flow rate of 3 litres per minute (L/min), and is capable of high temporal resolution PM mass concentration measurements. It can simultaneously measure PM_1, $PM_{2.5}$, PM_4, PM_{10}, and the total suspended PM. The instrument's measurement range is 1 to 150,000 µg m^{-3}. We used a factory-calibrated instrument for the monitoring campaign. On a daily basis, before starting measurements, the instrument was zero-calibrated and checked for any flow/filter errors. For more technical details on the instrument, refer to Viana et al. (2015), Rivas et al. (2017), and tsi.com.

3.1.3 microAeth

A microAeth (mA; model: AE51, Aethlabs, San Francisco, USA) was used to measure the on-road BC mass concentration. It is a palm-size, battery-powered instrument that measures the change in light attenuation at 880 nm with and without an aerosol sample and converts the attenuation measurements to BC mass concentration. The instrument collects air samples on a Teflon-coated glass fibre filter media, and the measurement range is from 1 to 1000 µg m^{-3}. It can be used at variable flow rates and logging intervals. For the current study, the instrument was configured to operate at 100 m L/min. The instrument is sensitive to vibration. More technical details about the instrument can be found at aethlabs.com.

3.1.4 Condensation Particle Counter

A condensation particle counter (CPC; model: 3007, TSI Incorporated, Minnesota, USA) was used to measure on-road fine particle number concentration. It is a battery-powered handheld portable instrument that can measure particles of size from 10 nm to >1 µm. The instrument

works on the principle of optical detection and has a flow rate of 0.7 L/min. Isopropyl alcohol is used as the working fluid. The alcohol-rich chamber in the instrument is supersaturated in alcohol, causing particles to grow in size, and the particles are then counted in the optical chamber. The instrument measurement range is from 0 to 100,000 cm^{-3}. More technical details of the CPC can be found at tsi.com. In the current study, CPC measurements were used to indicate levels of UFPs.

3.1.5 Ultrasonic Personal Air Sampler

An Ultrasonic Personal Air Sampler (UPAS; Access Sensor Technologies, Fort Collins, USA) was used for the filter-based sample collection of $PM_{2.5}$ during the mobile monitoring campaign. UPAS is a palm-sized battery-powered air monitor suitable for household and personal exposure monitoring. It consists of a 2.5-micron size selective inlet and operates at a flow rate of 1 L/min. A pre-weighed Teflon membrane filter (37 mm) is used for sampling the PM. The instrument can be connected to a proprietary mobile application through Bluetooth. More technical details on the UPAS monitor are available on accsensors.com.

3.1.6 Relative Humidity Monitor

Equinox (EQ-172), a palm-sized and battery-powered data logger, was used for on-road measurements of temperature and relative humidity. As the data logger can store a maximum of 16,350 data points, we used two loggers sequentially during monitoring hours. The measurement range of non-condensing relative humidity and temperature is from 0% to 100% and from –40 °C to +70 °C, respectively.

3.1.7 Mobile Platform

The mobile platform for the on-road air pollution measurements was a custom-fit CNG (compressed natural gas) car. Shelves for instruments were installed in the car by removing one of the rear passenger seats. All instruments were powered using their internal batteries. Instruments were placed close to the rear door windows of the car to ensure unobstructed airflow to them for sampling. Windows were kept open throughout the monitoring campaign. To reduce vibrations and ensure safety, instruments were given ample cushioning and secured using bungee cables.

3.2 Ambient Air Pollution Measuring Equipment

3.2.1 Beta Attenuation Monitor

A beta attenuation monitor (BAM; model: BAM-1022, Met One Instruments, Inc., Grants Pass, USA) was used to measure the ambient $PM_{2.5}$. The model used is a federal equivalent method (FEM) instrument certified by the United States Environmental Protection Agency (USEPA) for measuring 24-hour mean $PM_{2.5}$. The measurement technique of the instrument is based on the beta attenuation principle. It consists of a very sharp-cut 2.5 microns cyclone and a heated inlet that removes excess moisture in the sampled stream. BAM operates at a flow rate of 16.7 L/min. It has a ^{14}C isotope as the source for beta particles and a scintillation detector. The BAM is also equipped with meteorological sensors for measuring ambient temperature, relative humidity, and pressure. More details about the instrument are available at metone.com.

Figure 1: A collage of instruments used for on-road air pollution monitoring
(source: tsi.com, garmin.com; aethlabs.com; accsensors.com)

Figure 2: A collage of the instruments used for ambient air pollution monitoring (source: metone.com; aerosol.si)

3.2.2 Rack Mount Aethalometer

A Rack Mount Aethalometer (model: AE33, Aerosol Co. Ljubljana, Slovenia) was used to measure ambient BC mass concentrations. The instrument measures the optical absorption (same as that of mA) of an aerosol sample at seven wavelengths. The AE33 samples aerosol onto a glass fibre filter tape and analyses the attenuation of light through the sample. The measured attenuation is converted into absorbing aerosol mass concentration by the instrument software, following the Beer–Lambert–Bouguer law. The instrument has dual spot technology to compensate for loading errors. The absorbing mass concentration obtained at 880 nm is considered BC. We operated the AE33 at a flow rate of 2 L/min with a 2.5-micron cut cyclone attached at the end of the sample tubing. The AE33 can also apportion BC into fossil fuel combustion–emitted and biomass burning–emitted categories using an inbuilt algorithm. Detailed technical information of AE33 can be found at aerosol.si.

3.2.3 PurpleAir

For the LCS network, we used PurpleAir PA-II-SD (PurpleAir, Inc., USA) to measure ambient $PM_{2.5}$. PurpleAir (PA; Figure 3) is a low-cost PM monitor that uses a pair of PMS-5003 (Plantower Co., Ltd., China) laser counters to measure and log real-time PM_1, $PM_{2.5}$, and PM_{10}. In addition to PM measurements, PA also measures temperature, pressure, and relative humidity. PA is based on the Internet of Things (IoT) platform, and data are stored in a cloud server. PA also has an SD card version. The raw data from the instrument are logged at 2 minutes averaging intervals. The effective measurement range of $PM_{2.5}$ is from 0 to 500 µg m^{-3}. The maximum consistency error in $PM_{2.5}$ measurements is ±10% in the 100 to 500 µg m^{-3} range and ±10 µg m^{-3} in the 0 to 100 µg m^{-3} range. PA provides $PM_{2.5}$ data in two different channels, labelled $PM_{2.5}$ (cf_1) and $PM_{2.5}$ (cf_atm).

Figure 3: A PurpleAir LCS (source: purpleair.com)

3.3 Collocation Experiments

The DT and PA instruments measure $PM_{2.5}$ based on aerosol light scattering. As this technique is sensitive to various factors in addition to PM mass concentrations, correction is required for light scattering–based $PM_{2.5}$ measurements. To derive correction equation/factors, one of the popular methods is to collocate the devices with reference-grade instruments—federal equivalent method (FEM) or federal reference method (FRM)—and derive a statistical relationship between the two $PM_{2.5}$ measurement systems. As the composition of on-road and ambient $PM_{2.5}$ can be quite different, we conducted on-road collocation for DT and ambient collocation for PA. DT was collocated with the filter-based UPAS sampler during mobile monitoring rides (one filter sample for one day's ride). We collected 54 samples and executed eight blanks during the study period.

The pre-weighing and post-weighing of filters were carried out at the Colorado State University, Fort Collins. Details of the gravimetric analysis can be found in L'Orange et al. (2021). Based on this data, a study period mean correction factor for DT $PM_{2.5}$ was derived. Three PA sensors were collocated with BAM on the CSTEP roof terrace throughout the study period. Real-time data from both instruments were then used to derive month-wise correction equations for PA $PM_{2.5}$ measurements.

For the performance evaluation of various LCSs in measuring $PM_{2.5}$, we collocated 10 different LCSs with BAM throughout the study period. The sensors under evaluation were mostly assembled in India.

3.4 Data Cleaning and Correction

Because of the portable and semi-reference-grade nature and limited dynamic range of instruments, the air pollution data collected using such instruments often require a quality check and subsequent corrections. Most mobile monitoring instruments needed instrument-specific corrections.

3.4.1 Global Positioning System Measurements

Owing to uncertainties in GPS measurements, latitude and longitude information from the Garmin device was required to snap to the nearest road feature for mobile monitoring measurements in most cases. We considered a maximum of 30-metre error distance for the snapping procedure, beyond which data were discarded. We used the *k*-nearest neighbour algorithm to snap measurements to the nearest OpenStreetMap (OSM) road feature. The 'snapPointsToLines' function in the 'R' programming language was used for snapping.

3.4.2 DustTrak $PM_{2.5}$

The correction factors obtained from the DustTrak and UPAS collocation experiments were applied to 1 Hz on-road $PM_{2.5}$ data collected during the mobile monitoring campaign.

3.4.3 microAeth BC

As the mA instrument is sensitive to vibration, which happens during mobile measurements, we used the algorithm developed by Apte et al. (2011) to remove spurious BC data. Next, we applied a loading correction algorithm demonstrated in Ban-Weiss et al. (2009).

3.4.4 Condensation Particle Counter UFPs

The CPC's dynamic range of particle number concentration measurements is less than that of Bengaluru on-road levels. To overcome this, we used the CPC-3007 along with a diluter (characterised by a dilution ratio of ~5.5; Ban-Weiss et al., 2009). The measurements were corrected to the dilution ratio for further analysis.

3.4.5 Hour-of-the-Day Correction

Pollution levels vary both spatially and temporally. Although mobile measurements capture spatial variation, we need to correct temporal bias to account for diurnal variations in emissions and boundary layer height. To do this, we applied an hour-of-the-day correction (Apte et al., 2017) in addition to the instrument-specific corrections described above. This method uses multiplicative correction factors derived using ambient measurements from BAM for $PM_{2.5}$ and AE33 measurements for BC. Given that ambient measurements were not available for UFPs, we

applied BC correction factors to UFPs measurements. Month-wise factors were derived and applied to on-road air pollution data.

3.4.6 PurpleAir $PM_{2.5}$

$PM_{2.5}$ data from individual dual laser counters were temporally averaged to one-hour intervals. A quality check criterion was applied on the hourly PA $PM_{2.5}$ based on the difference in values from the dual Plantower laser counters, following Barkjohn et al. (2021). $PM_{2.5}$ is considered valid only if the difference between PA's measurements from the dual laser counters is less than 5 µg m^{-3} or 61% of the absolute concentration reading. Correction factors obtained from PA and BAM collocation experiments were used to correct the one-hour averaged $PM_{2.5}$ from PA. Using corrected hourly $PM_{2.5}$ measurements from PA, daily, monthly, and seasonal averages were computed, following 75% completeness criteria.

3.4.7 Gridding

The on-road 1 Hz pollution data was gridded to 50 m uniform road segments. On a daily basis, all 1 Hz measurements falling in the grid were averaged. Using daily gridded data, a grid-wise median of daily means was computed to represent a particular grid pollution level.

Because of the gold standard nature of BAM and AE33, no corrections were applied to their measurements. More details on the correction algorithms mentioned above can be found in CSTEP & ILK Labs (2020). A schematic of all the data handling and correction procedures is given in Figure 4.

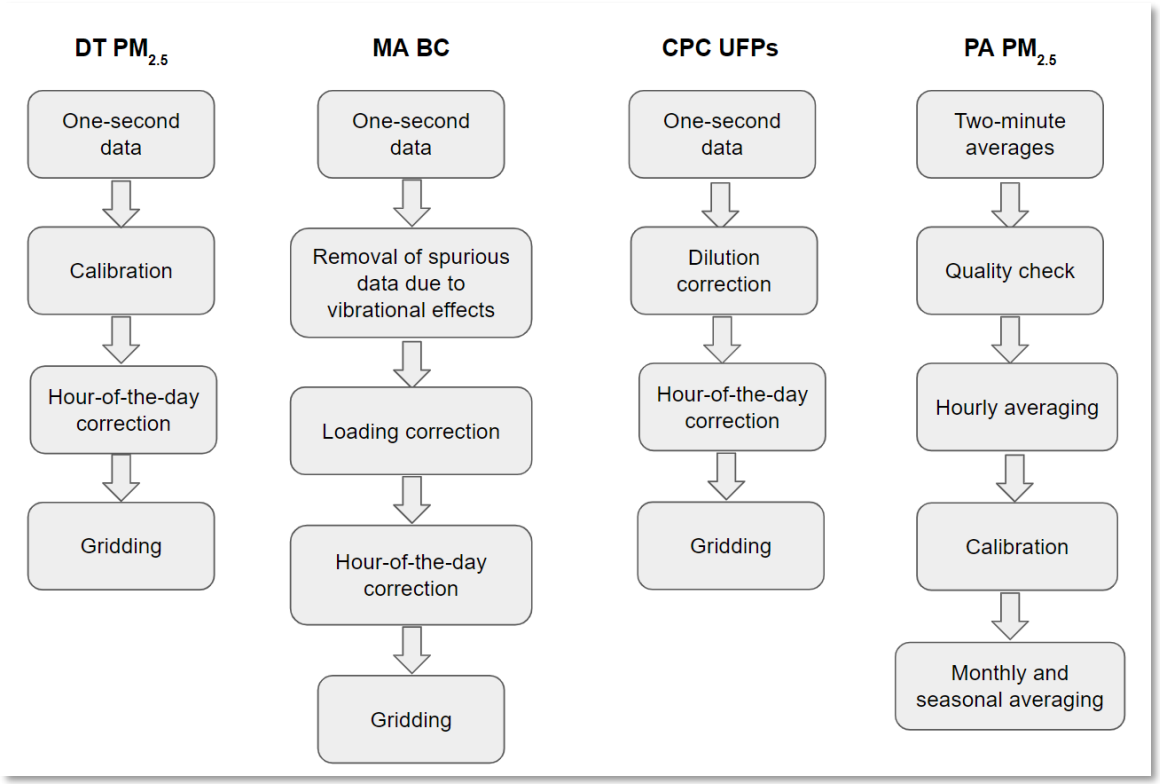


Figure 4: Schematic illustrating various cleaning and correction procedures applied to collected data

3.5 PurpleAir Network

We established a city-wide PA network to measure ambient $PM_{2.5}$. The installation was carried out following best practices outlined in CSTEP & ILK Labs (2022a). Under this network, around 55 sensors were installed, covering various land-use categories in the Bengaluru Rural and Urban districts. Out of the 55 sensors, 8 sensors were installed in the Bengaluru Rural district, 5 in peri-urban Bengaluru, and the rest in the Bruhat Bengaluru Mahanagara Palike (BBMP; Bengaluru municipal corporation jurisdiction) area. All sensors deployed in the field were new or refurbished. The network was set up / refurbished between October and November 2021. In addition to using ambient $PM_{2.5}$ measurements from the PA network, we also used reference-grade $PM_{2.5}$ data from eight continuous ambient air quality monitoring stations (CAAQMS) installed by the Karnataka State and Central Pollution Control Boards (PCB). Geographical locations of the PA sensors and CAAQMS are shown in Figure 5.

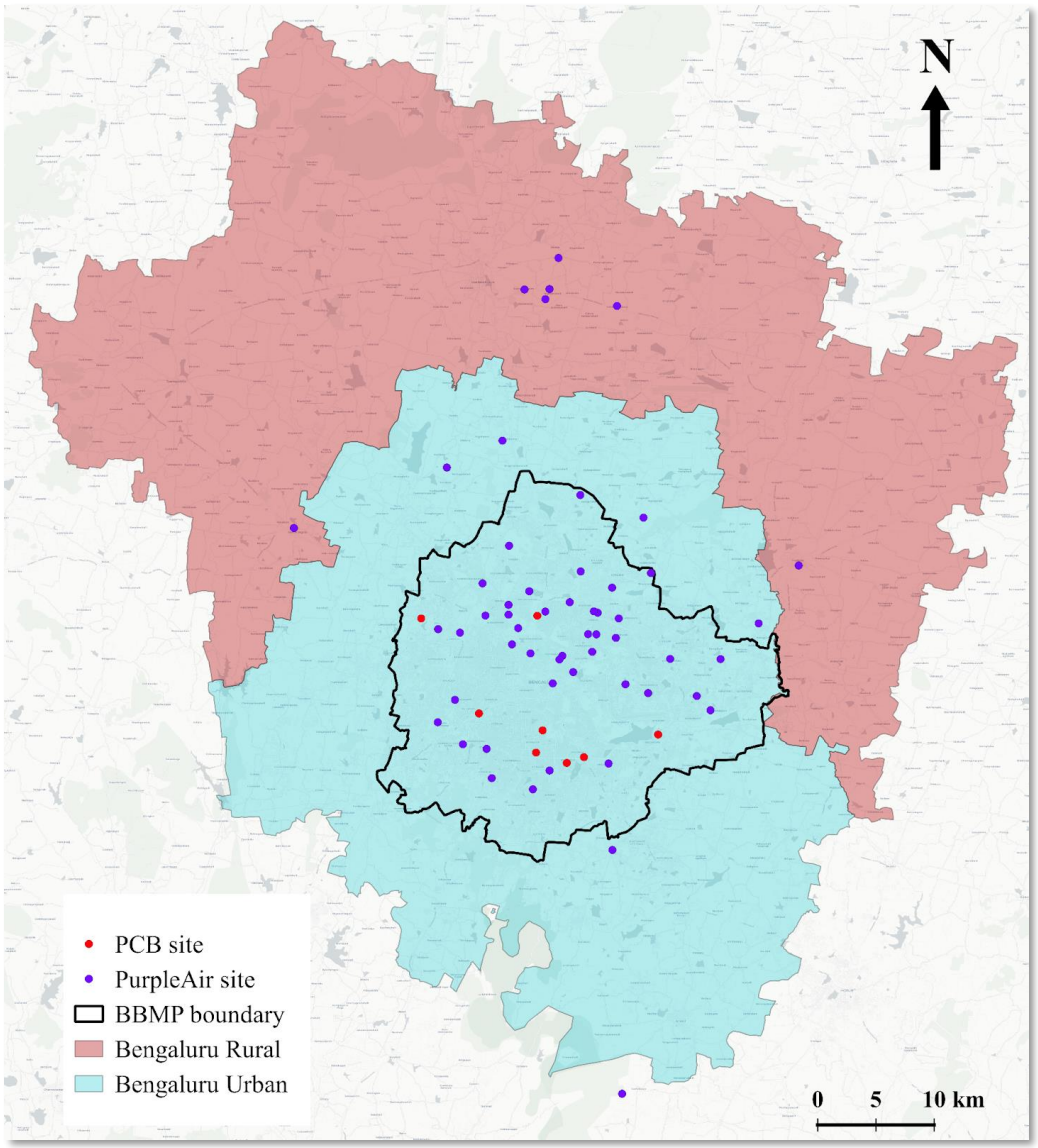


Figure 5: Geographical locations of the PA network and the PCB (regulatory) stations

3.6 Mobile Monitoring Routes

In this study, we aimed to measure on-road air pollution levels of the Bengaluru Urban district, covering around 10% of the total road network (i.e., 10% of ~11,500 km). Around 1,200 km of roadways (~900 unique km) were identified for the mobile monitoring campaign, and they were divided into 11 unique routes (each route covering around 100 km). Each route was covered within a day during daytime (from 9 a.m. to 5 p.m.). The majority of the monitoring happened between December 2021 and May 2022, but a few rides were conducted in the months of November 2021 and June 2022. Most of the rides were conducted on non-rainy weekdays. Each route was also mapped on weekends at least once during the study period. All unique routes (11) were driven at least four times, three routes were driven eight times, and five routes were driven 12 times. Based on the OSM road classification, routes were categorised into four types—major, arterial, residential, and unclassified roads—by combining relevant categories. The study routes are shown in Figure 6, and more details of unique routes (as abbreviated in the bottom panel of Figure 6) are provided in Table A1. The total distance driven during the sampling campaign was ~10,600 km (via 665 drive hours), and more than 2 million data points were collected.

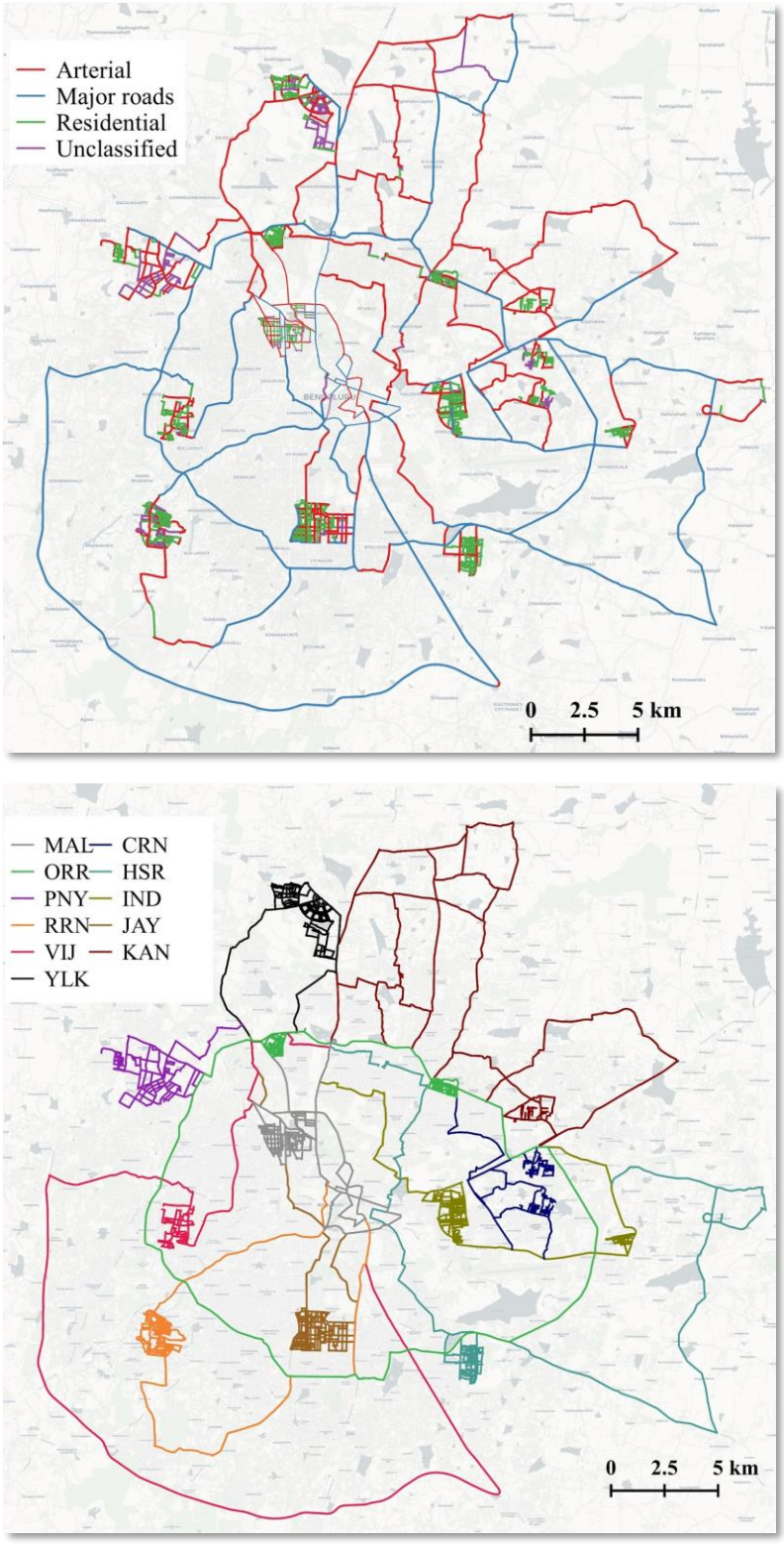


Figure 6: Mobile monitoring route (top panel) colour indicates the type of the road; colour (bottom panel) indicates the unique route

3.7 Land-Use Regression

Land-use regression (LUR) is a well-established model for simulating high-resolution spatial variability of urban-level air pollution (Hoek et al., 2008; Eeftens et al., 2012; Messier, 2018). In this study, we adopted the LUR methodology developed in the European Study of Cohorts for Air Pollution Effects (ESCAPE) study. The methodology was built on a supervised stepwise regression procedure. As the first step, mean pollutant concentrations were regressed linearly with shortlisted predictors, and the predictor giving the highest adjusted coefficient of determination (R^2) was included in the model if the direction of effect was as defined *a priori*. A positive (negative) direction of effect indicated that the predictor acted as a source (sink) of air pollution. In the next step, each predictor was added to the model one by one, and the predictor resulting in the highest gain in the model adjusted-R^2 and following the direction of effect was retained. The procedure was continued until the addition of any predictor to the model resulted in a gain of less than 0.01 in the model adjusted-R^2. As the final steps of the model training, the potential predictors remaining in the model were tested for their statistical significance (p-values) and collinearity or variance inflation factor.

We sequentially removed all predictors with p-values greater than 0.1. The influential observations (pollutants) were identified based on Cook's D values. The model was reassessed after removing non-significant and collinear predictors and influential observations. The model's performance was evaluated based on a 10-fold cross-validation procedure. In the 10-fold CV method, we randomly split the spatial data into 10 folds and fitted the model using 9 folds and predict the remaining fold, with the predictions evaluated against the true value. This process was repeated ten times, thus ensuring every fold was evaluated. The final step was to calculate Moran's I value to investigate the spatial autocorrelation in model residuals. In this study, we trained LURs using the monthly and seasonal ambient $PM_{2.5}$ (using data collected by PurpleAir network sensors) and on-road BC, $PM_{2.5}$, and UFPs (using data sets collected by mobile monitoring). Once the model was trained, spatial predictions of pollutant levels were carried at 50 m resolution.

3.7.1 Predictor Variables

Predictor variables are selected based on their linkage to an area's environmental characteristics, especially those that influence pollution emission intensity and dispersion efficiency. The list of predictors considered in the current study, their buffer sizes, and their direction of effect are listed in Table 1. The broad classification of predictors includes (i) land use and land cover (LULC), (ii) demographic and geographic variables, (iii) road and rail network, (iv) satellite-based pollution parameters, and (v) reanalysis-based meteorological parameters. The buffer sizes of eligible predictors were adopted from Eeftens et al. (2012). For model training, we excluded predictors where most (>75%) values were the same.

Table 1: The list of shortlisted predictors used for LUR model training

Data set (source)	Predictor	Unit	Buffer size (m)	Direction of effect
Land use (ESA Worldcover, 2020 product)	Surface area of built-up area	m^2	100, 300, 500, 1000, 5000	+
	Surface area of permanent water bodies	m^2	100, 300, 500, 1000, 5000	-
	Surface area of tree cover	m^2	100, 300, 500, 1000, 5000	-
	Surface area of bare/sparse vegetation	m^2	100, 300, 500, 1000, 5000	+/-
	Surface area of cropland	m^2	100, 300, 500, 1000, 5000	+/-
	Sum of grassland, scrubland/shrubland, and herbaceous wetland	m^2	100, 300, 500, 1000, 5000	-
Population (GHSL, 2015 product)	Number of inhabitants	#	300, 500, 1000, 5000	+
Elevation (SRTM)	Square root of the elevation	$m^{1/2}$	NA	-
Sensor height	Square root of the sensor height	$m^{1/2}$	NA	-
Geographical variables (GPS)	Longitude	^oE	NA	+/-
	Latitude	^oN	NA	+/-
Road network (OSM, 2021 product)	Road length of total roads in buffer	m	25, 50, 100, 300, 500, 1000	+
	Road length of residential roads in buffer	m	25, 50, 100, 300, 500, 1000	+
	Road length of primary and secondary roads	m	25, 50, 100, 300, 500, 1000	+
	Road length of tertiary roads	m	25, 50, 100, 300, 500, 1000	+
	Inverse distance and inverse square distance to the nearest road	m^{-1}; m^{-2}	NA	+
Train network (OSM)	Length of train tracks within the buffer	m	100, 300, 500, 1000, 5000	+
Airport (OSM)	Inverse distance to the airport	m^{-1}	NA	+
Industry sources (secondary data)	Inverse distance to the nearest industry	m^{-1}	NA	+

Data set (source)	Predictor	Unit	Buffer size (m)	Direction of effect
NDVI (Sentinel 2)	Surface area of green vegetation	-1 to +1	100, 300, 500, 1000, 5000	-
Light intensity (VIIRS)	Night-time light intensity		NA	+
NO_2 (TROPOMI)	Tropospheric vertical column of NO_2	mole m^{-2}	NA	+
AOD (MODIS-MAIAC)	Aerosol optical depth	-	NA	+
Meteorological data (ERA5-Land)	2 m temperature of air	K	NA	-
	2 m pressure	Pa	NA	+
	10 m eastward component of wind	m/s	NA	+/-
	10 m northward component of wind	m/s	NA	+/-
	10 m wind speed	m/s	NA	-
	Total precipitation	m	NA	-

ESA: European Space Agency (Zanaga et al., 2021); GHSL: Global Human Settlement Layer (Freire et al., 2016); SRTM: Shuttle Radar Topography Mission (Farr et al., 2007); OSM: OpenStreetMap (OpenStreetMap contributors, 2021); NDVI: Normalised Difference Vegetation Index (Gascon et al., 2016); VIIRS: Visible Infrared Imaging Radiometer Suite (Veefkind et al., 2012); TROPOMI: TROPOspheric Monitoring Instrument (Van Geffen et al., 2020); MODIS-MAIAC: MODerate Resolution Imaging Spectroradiometer-Multi-Angle Implementation of Atmospheric Correction (Lyapustin et al., 2018); ERA: European Centre for Medium-Range Weather Forecasts Reanalysis (Muñoz-Sabater et al., 2021).

3.8 Performance metrics

We used the following statistical metrics to evaluate the performance of the models trained. They are (i) coefficient of determination (R^2), (ii) root mean square error (RMSE), and (iii) normalised root mean square error (NRMSE):

$$R^2 = \left(\frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right)^2} \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} y_i\right)^2}} \right)^2 \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}} \quad (2)$$

$$NRMSE = \frac{RMSE}{\frac{1}{n} \sum_{i=1}^{n} y_i} \quad (3)$$

where x and y represent pollutant concentrations from two different data sets and n is the number of data points. For the NRMSE calculation, we used the gold standard data set (labelled y) mean to normalise RMSE.

4. Results

4.1 Performance Evaluation of Various LCSs

Preliminary results of evaluations have already been reported in CSTEP and ILK Labs (2022b). We present the seasonal comparison of the performance of LCSs here. The performance of 10 PM LCSs was compared. They are PurpleAir, Airveda, PAQS, BlueSky, Atmos I, Atmos II, Aerogram, Prkruti, Aurassure, and Prana Air (Figure 7). Tables A2 and A3 summarise several physical, technical, and operational characteristics of the LCSs under evaluation. For deriving performance metrics, individual sensors' $PM_{2.5}$ data (characterised by different logging and/or averaging periods) were temporally averaged to one hour and daily (24 h) periods.

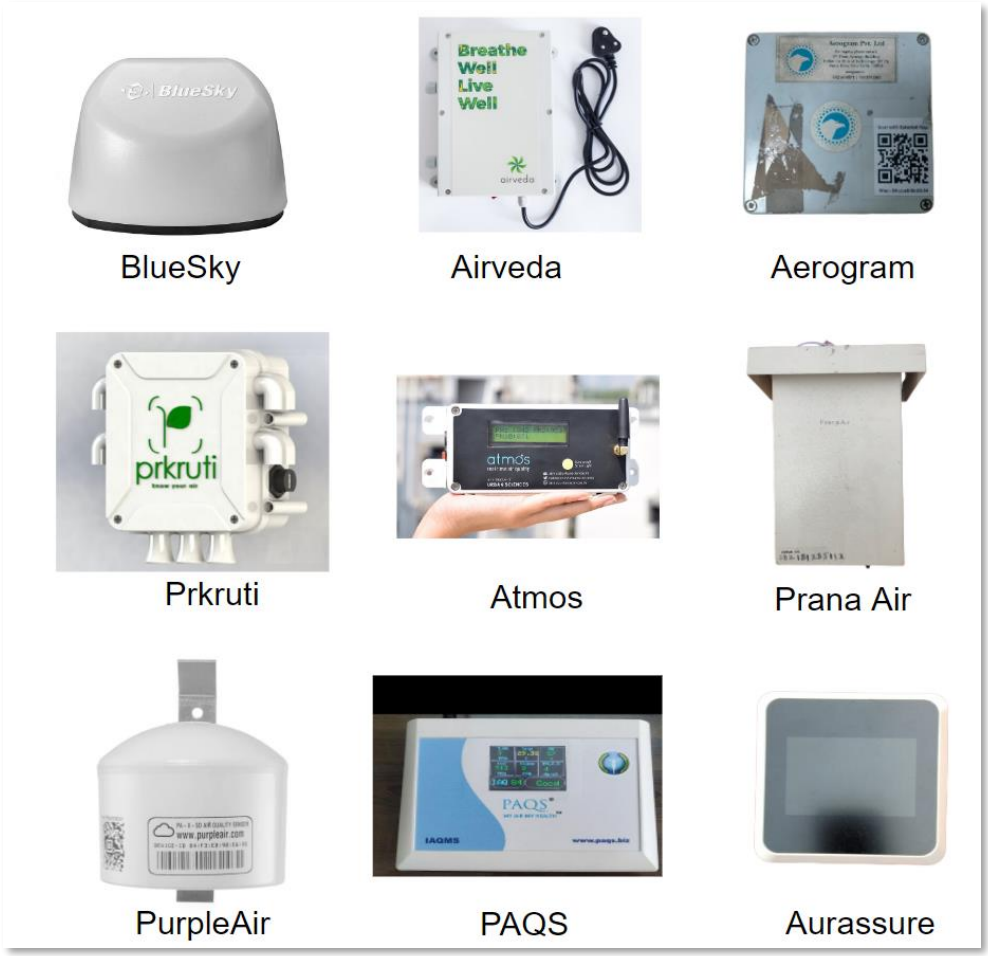


Figure 7: Low-cost sensors (LCSs) under evaluation

Figure A1 shows the time series of the daily mean $PM_{2.5}$ for the 10 LCSs and BAM. A few LCSs overestimated $PM_{2.5}$ compared to BAM measurements, and a few underestimated. Overall, the LCSs $PM_{2.5}$ measurements captured the temporal trend in $PM_{2.5}$. The seasonal mean (standard deviation) BAM $PM_{2.5}$ measurements were ~44 µg m^{-3} (15.2 µg m^{-3}) and 34 µg m^{-3} (10.7 µg m^{-3}) for winter (DJF) and summer (MAM), respectively. Figures 8 and 9 show scatter plots of hourly mean $PM_{2.5}$ from BAM and LCSs for winter and summer, respectively. It can be observed that the bias in LCS $PM_{2.5}$ is not uniform across sensors. Table 2 summarises the season-wise performance metrics of the LCSs $PM_{2.5}$. Across LCSs, no remarkable seasonal differences in performance were observed. The highest bias was observed in PAQS $PM_{2.5}$, with RMSE values of ~34.1 µg m^{-3} and 33.2 µg m^{-3} for winter and summer, respectively. NRMSE values ranged between 0.22 and 0.80 for

winter and 0.23 and 0.86 for summer. The performance of PurpleAir, Airveda, Prana Air, BlueSky, and Prkruti was almost similar across seasons. Aerogram exhibited better performance in summer compared to winter. $PM_{2.5}$ from Aurassure was evaluated for the summer season only as it was installed late.

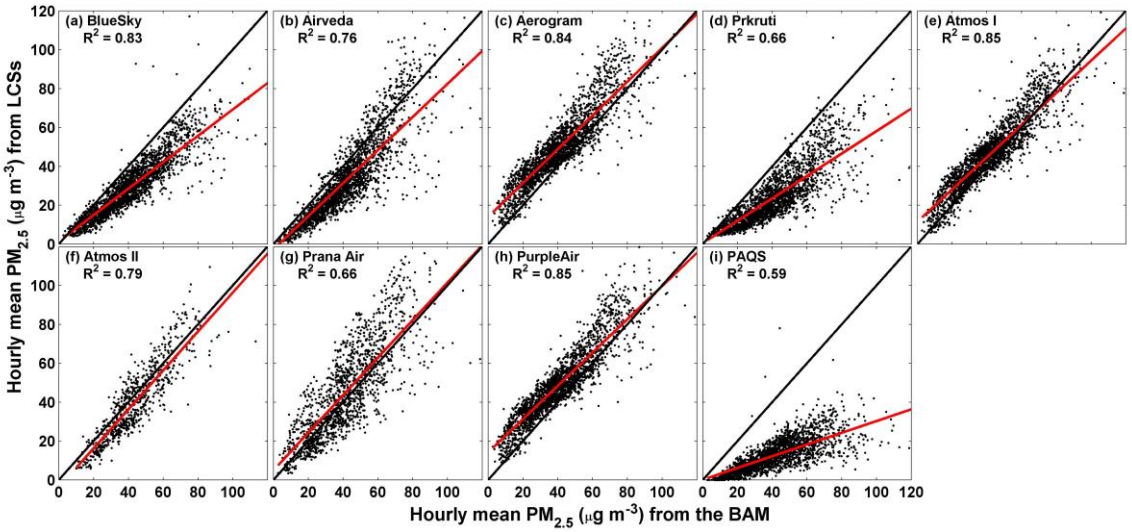


Figure 8: Scatter plots of hourly mean LCSs and BAM $PM_{2.5}$ for winter (DJF)

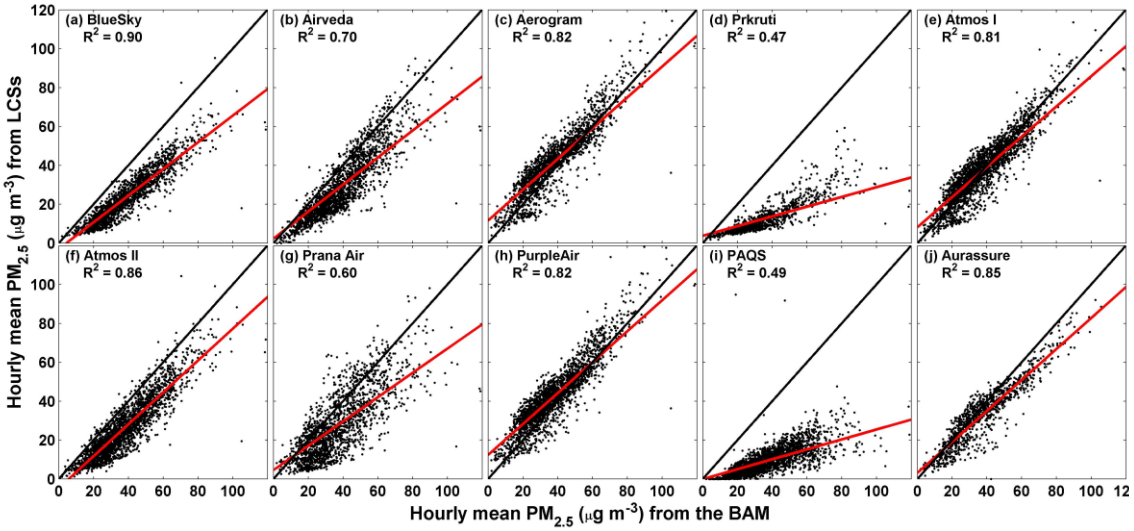


Figure 9: Scatter plots of hourly mean LCSs and BAM $PM_{2.5}$ for summer (MAM)

Table 2: Season-wise performance statistics of hourly LCS $PM_{2.5}$ measurements
(N represents the number of paired data points)

Sensor	N		R^2		RMSE (µg m^{-3})		NRMSE	
	DJF	MAM	DJF	MAM	DJF	MAM	DJF	MAM
BlueSky	2108	1719	0.83	0.90	15.9	17.9	0.37	0.44
Airveda	2111	1859	0.76	0.70	14.4	15.3	0.34	0.38
Aerogram	2092	1747	0.84	0.82	12.5	9.8	0.29	0.25
Prkruti	2058	1569	0.66	0.46	21.4	28.2	0.51	0.77
Atmos I	2111	2140	0.85	0.81	9.9	9.0	0.23	0.23
Atmos II	684	2140	0.79	0.86	9.5	13.9	0.22	0.36
Prana Air	1371	1727	0.66	0.60	14.4	16.9	0.34	0.44
PurpleAir (cf_atm)	2092	2046	0.85	0.82	12.0	9.9	0.28	0.25
PAQS	2105	2122	0.59	0.49	34.1	33.2	0.80	0.86
Aurassure	NA	1008	NA	0.85	NA	8.5	NA	0.23

Further, LCS performance was evaluated based on daily mean $PM_{2.5}$. Figures A2 and A3 show scatter plots of daily $PM_{2.5}$ from BAM and LCSs for winter and summer, respectively. The performance metrics indicate that the daily mean LCS $PM_{2.5}$ is more accurate and precise compared to the hourly averages (Table A4). The NRMSE of the daily mean LCS $PM_{2.5}$ varied between 0.11 and 0.74 for winter and 0.11 and 0.78 for summer.

In addition to $PM_{2.5}$, we also evaluated the performance of LCSs (Figures A4 and A5 and Table A6) in measuring ambient relative humidity (RH) and temperature (T). We used the meteorological station–measured RH and T as gold standard measurements. Most of the LCSs overestimated T and underestimated RH (except Prana Air). Prana Air measurements of RH were relatively accurate (R^2=0.93, RMSE=6.7%, and NRMSE=0.11), and PAQS measurements of T were relatively accurate (R^2=0.87, RMSE=3.0 ^{0}C, and NRMSE=0.12). The performance evaluation of RH and T measured by LCSs is vital as these parameters can be used in developing correction equations to improve the accuracy of LCS $PM_{2.5}$ measurements.

4.2 Calibration of Optical $PM_{2.5}$

4.2.1 DT $PM_{2.5}$

We used on-road gravimetric $PM_{2.5}$ to correct on-road 1 Hz DT $PM_{2.5}$. We estimated the limit of quantitation (LOQ) for the filter sample dust mass using the field blank dust mass values. Mean plus five times the standard deviation of the field blank dust mass values were used as the LOQ metric, which was found to be ~50 µg. The sample mass below the LOQ was removed from further estimation of gravimetric $PM_{2.5}$. Out of 54 filter samples, only 19 samples qualified for the LOQ condition. The ratio of averaged gravimetric $PM_{2.5}$ (combining all 19 samples) to averaged DT $PM_{2.5}$ was estimated and used as the correction factor to correct on-road DT $PM_{2.5}$. The correction factor (CF) was found to be ~0.56 ± 0.27. As data were collected during daytime, we have not applied any RH correction factors to DT $PM_{2.5}$.

$$CF_{DT-PM_{2.5}} = \frac{PM_{2.5-gravimetric}}{PM_{2.5-DT}} \quad (4)$$

4.2.2 PA $PM_{2.5}$

Month-wise bias (RMSE) of uncorrected hourly PA $PM_{2.5}$ ranged between ~13 and 35 µg m^{-3}. To improve the accuracy of PA $PM_{2.5}$, we explored training various statistical regression models—which include univariate, multivariate, linear mixed-effects, and general additive models—using collocation data to arrive at the best-performing PA correction model for Bengaluru. The model training was performed using hourly BAM $PM_{2.5}$, PA $PM_{2.5}$, PA T, and PA RH. A multivariate regression model (with $PM_{2.5}$ and RH terms) for cf_1 channel PA $PM_{2.5}$ emerged as the best performing model. This is in line with the observation made by Barkjohn et al. (2021).

We built month-wise models to account for any temporality in model coefficients. Regression coefficients of trained models for individual months are provided in Table A7. All the coefficients are statistically significant. Scatter plots between model-corrected PA $PM_{2.5}$ and BAM $PM_{2.5}$ are shown in Figure 10. Month-wise performance metrics of the corrected PA $PM_{2.5}$ are provided in Table A8. After the application of the correction factors, the accuracy of corrected PA $PM_{2.5}$ improved significantly, and the month-wise bias (RMSE) of hourly PA $PM_{2.5}$ ranged between 5 and 10 µg m^{-3}. The accuracy in the corrected PA $PM_{2.5}$ further improved when hourly corrected PA $PM_{2.5}$ was averaged to daily intervals.

For further analysis and LUR model building, we used only the corrected DT $PM_{2.5}$ and PA $PM_{2.5}$.

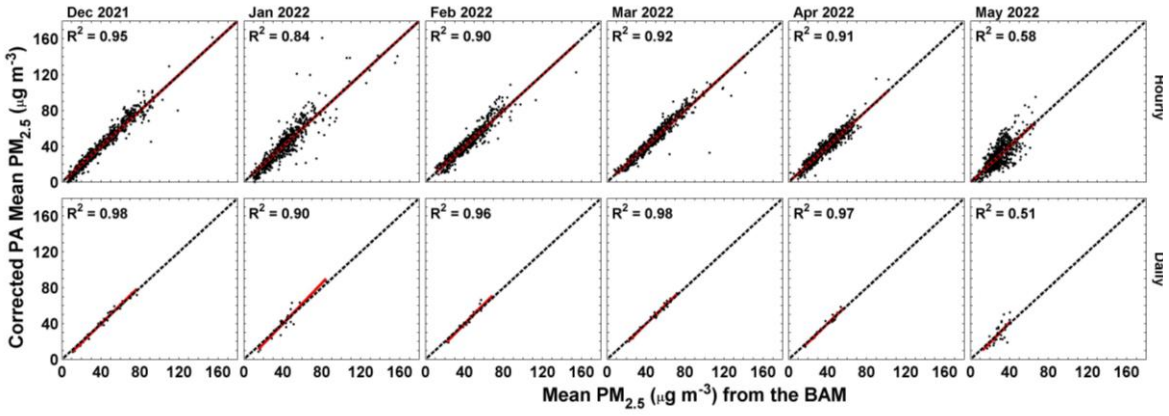


Figure 10: Month-wise scatter plots between corrected hourly PA $PM_{2.5}$ and BAM $PM_{2.5}$ (top row) and daily mean corrected PA $PM_{2.5}$ and daily mean BAM $PM_{2.5}$ (bottom row)

4.3 Mobile Monitoring 'Data-Only' Pollution Route Maps and Statistics

The ride-to-ride (all unique routes combined) variations in the gridded pollutant levels are shown in Figure A6. From the figure, it is clear that spatial variability (irrespective of the route) is higher in BC and UFPs compared to the variability in $PM_{2.5}$. Following our previous study (CSTEP and ILK labs, 2020), we used the median of gridded daily mean pollutant levels to generate route maps of on-road air pollution, representing the study period's central tendency. Figure 11 shows the route map of on-road $PM_{2.5}$, BC, and UFPs generated from four repeat rides. Quantile colour scales are used. Across the study route, on-road $PM_{2.5}$ ranged from ~13 µg m^{-3} to >1000 µg m^{-3}, BC ranged from ~1 µg m^{-3} to >2000 µg m^{-3}, and UFPs ranged from ~2 x 10^3 cm^{-3} to 420 x 10^3 cm^{-3}. Spatially averaged pollution levels after four passes were ~88 µg m^{-3} for $PM_{2.5}$, 36 µg m^{-3} for BC, and ~67 x 10^3 cm^{-3} for UFPs. Route maps from eight and 12 repeat rides are provided in Figure A7 and Figure 12, respectively.

Not surprisingly, major roadways are characterised by the highest levels of pollution. Wall-to-wall driving, where the vehicle is driven through most of the roads in the neighbourhood, showed the highest levels in the industrial area of Peenya (PNY). The lowest $PM_{2.5}$ concentrations were observed in the Indiranagar (IND) area, a residential neighbourhood, and similar results were seen even after 12 passes (Figure 12). Residential areas (Indiranagar, HSR Layout, Rajarajeshwari Nagar, and Malleswaram) clearly showed a gradient; pollution levels decreased as we moved away from the main road. BC and UFP spatial maps also exhibited a spatial pattern similar to $PM_{2.5}$.

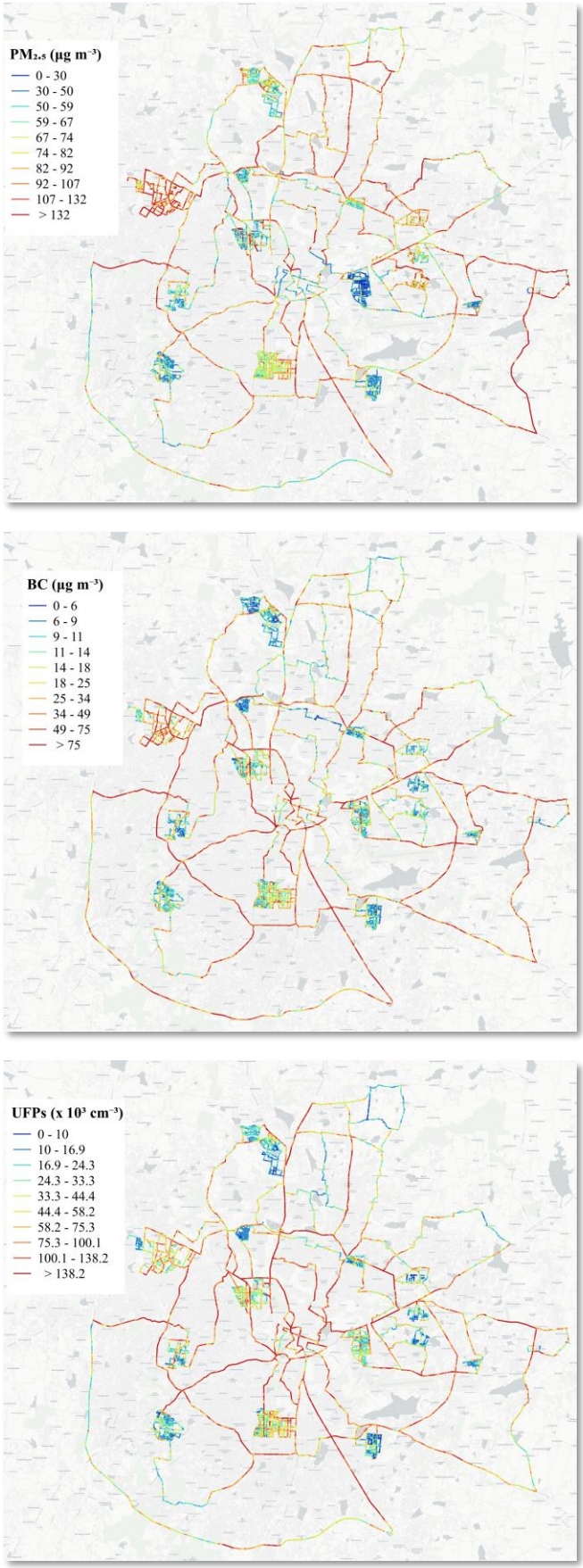


Figure 11: Route maps of on-road $PM_{2.5}$, BC, and UFPs derived using data from four repeat rides

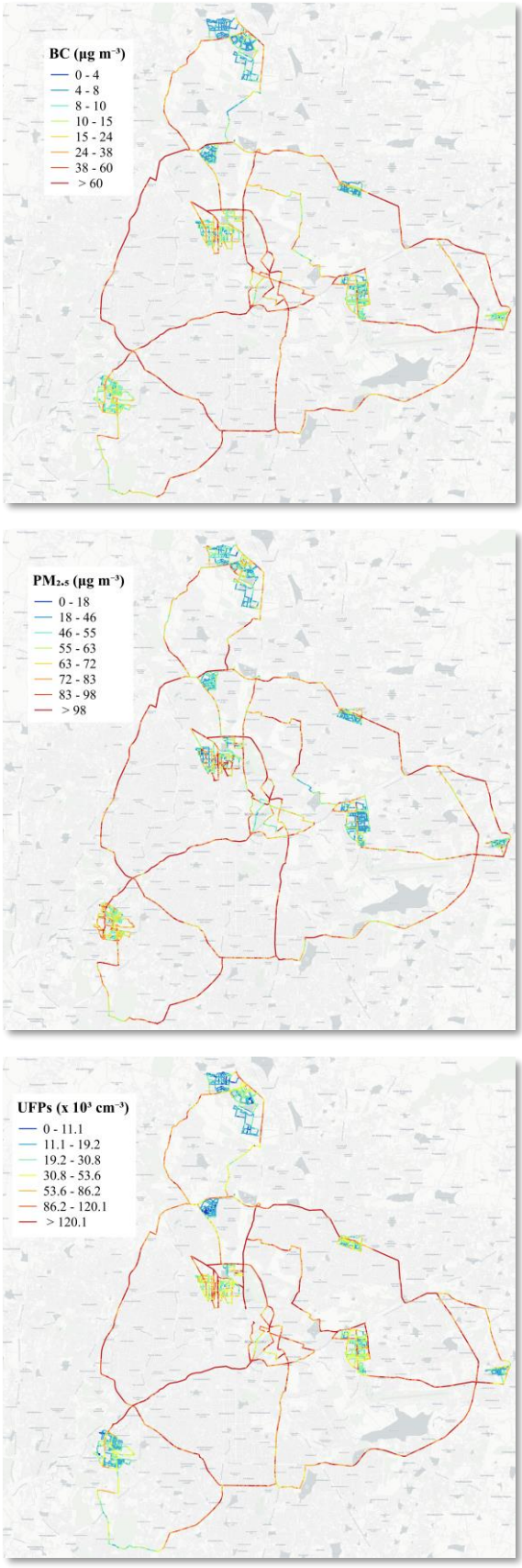


Figure 12: Route maps of on-road $PM_{2.5}$, BC, and UFPs derived using data from 12 repeat rides

When grouped by road classification (Figure 13), residential roads experienced lower concentrations, followed by arterial and major roads. The difference between major and arterial roads was more pronounced for BC and UFPs. In our study area, we had 81 km of road that was unclassified (as per OSM). Unclassified roads are shown separately in the plots. The average pollutant levels for unclassified road segments were found to be higher than the levels for residential roads. See Table 3 for more detailed statistics.

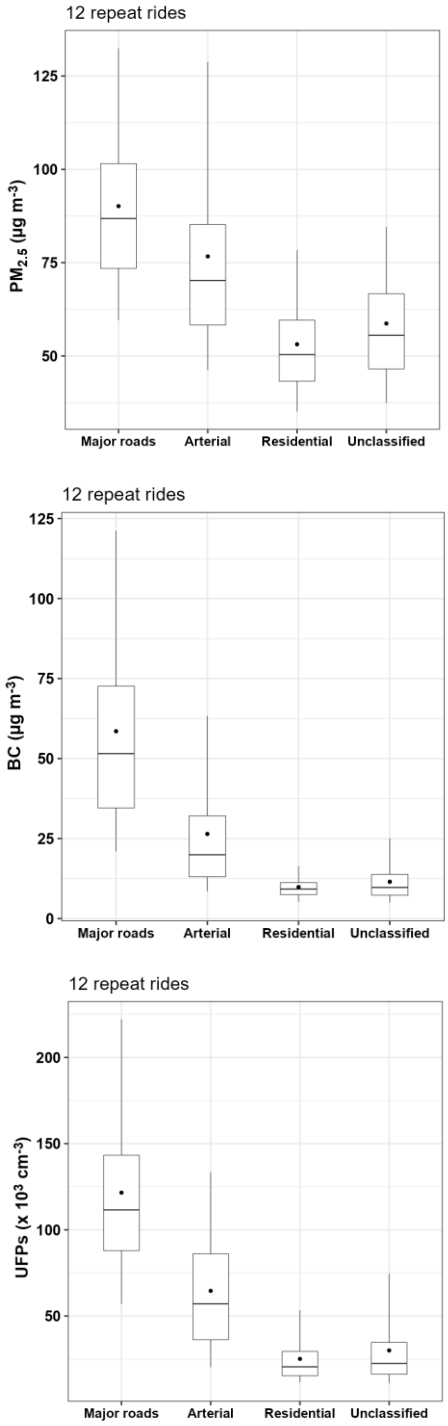


Figure 13: Distribution of on-road pollutant levels by road type. The central line of the box represents the median, the dot represents the mean, the box represents the inter-quartile range (IQR), and whiskers represent the 5th and 95th percentiles. Outliers are omitted for readability.

Table 3: Road classification–wise pollutant level statistics. SD represents the standard deviation

12 repeat rides road type	Mean	SD	5th percentile	25th percentile	50th percentile	75th percentile	95th percentile	Minimum	Maximum
$PM_{2.5}$ ($\mu g\ m^{-3}$)									
Major	90	23	60	73	87	102	132	45	222
Arterial	77	31	46	58	70	85	129	25	372
Residential	53	17	35	43	50	60	78	13	264
Unclassified	59	25	37	47	56	67	85	22	511
All roads combined	73	29	40	53	67	86	122	13	511
BC ($\mu g\ m^{-3}$)									
Major	59	33	21	35	52	73	121	10	289
Arterial	26	22	9	13	20	32	63	4	227
Residential	10	4	5	7	9	11	16	2	44
Unclassified	12	7	5	7	10	14	25	3	52
All roads combined	31	30	6	10	18	43	91	2	289
UFPs (x 10^3 cm^{-3})									
Major	122	50	57	88	112	143	222	20	340
Arterial	65	37	20	36	57	86	133	10	271
Residential	25	16	12	15	21	30	53	6	167
Unclassified	30	23	11	16	23	35	75	7	184
All roads combined	68	54	13	24	52	101	171	6	340

Based on the common route data, we compared pollution levels from four and eight repeat rides with levels from 12 repeat rides. The NRMSE (with respect to central tendencies of the 12 repeat rides) of the four repeat routes' $PM_{2.5}$, BC, and UFPs is ~0.35, 0.65, and 0.40, respectively. Similarly, the NRMSE of the eight repeat routes' $PM_{2.5}$, BC, and UFPs is ~0.20, 0.27, and 0.20, respectively. This clearly shows that spatial representativeness of pollution levels increases with an increase in the number of repeat rides.

On-road concentrations of $PM_{2.5}$ and BC were consistently higher than ambient levels throughout the study period (Figure A8). On-road $PM_{2.5}$ was around two to three times higher than ambient $PM_{2.5}$ levels, while on-road BC was around five to 10 times higher than ambient BC levels. Here, ambient levels refer to reference-grade measurements carried out at CSTEP.

4.4 Ambient $PM_{2.5}$ Statistics

The precision of PA $PM_{2.5}$ was investigated in several in-house collocation experiments (before field deployment) and found to be satisfactory. The coefficient of variation was well within the target metrics specified by USEPA (Duvall et al., 2021). The accuracy in PA $PM_{2.5}$ measurements significantly improved after they were corrected using the month-specific correction equation. Box plots in Figure A9 depict spatio-temporal variations in the daily mean $PM_{2.5}$. The plot includes corrected $PM_{2.5}$ data from the PA network and regulatory $PM_{2.5}$ measurements from PCBs. The highest $PM_{2.5}$ values were observed during March 2022, followed by December 2021, February 2022, January 2022, April 2022, and May 2022. See Table 4 for month-wise site mean statistics. The $PM_{2.5}$ seasonal mean across all sites was 42 µg m^{-3} and 36 µg m^{-3} for winter and summer, respectively. The spatial variability was higher during winter months (as evident from the range, SD, and IQR) compared to summer months. Ambient $PM_{2.5}$ levels in peri-urban and rural areas of Bengaluru were comparable with urban $PM_{2.5}$ levels.

Table 4: Month-wise site mean statistics

Pollutant	Mean	SD	5th percentile	25th percentile	50th percentile	75th percentile	95th percentile	Minimum	Maximum
December 2021	45	5.4	36	41	45	48	54	33	57
January 2022	39	6.1	30	35	39	42	50	22	55
February 2022	43	5.5	34	41	44	46	51	27	60
March 2022	48	5.5	38	44	48	52	58	37	61
April 2022	35	4.5	27	33	35	37	42	22	48
May 2022	26	4.9	19	23	27	29	35	18	40

4.5 Land-Use Regression Modelling

4.5.1 On-road Air Pollution Models

Separate LUR models were developed for on-road $PM_{2.5}$, BC, and UFPs using data from the mobile monitoring of roads with four, eight, and 12 repeat rides. Thus, nine distinct LURs were developed (three pollutants and three levels of repeat rides).

The 'four repeat rides' data set (data from all roads with four repeat rides) represents the largest road length: 877 km, 308 drive hours, and ~17k road segments. The 'eight repeat rides' data set (data from all roads with eight repeat rides) represents a relatively smaller road length: 686 km, 448 drive hours, and ~13k road segments. The '12 repeat rides' data set (data from all roads with twelve repeat rides) comprises the smallest road length of the three: 421 km, 675 drive hours, and ~8k road segments.

Table 5 presents the results of the nine LUR models. For the mobile monitoring models, we used only the daytime mean meteorological parameters. From the table, it is apparent that LUR models using data from 12 repeat rides performed better compared to others (models using data from four and eight repeat rides) for all the pollutants. For instance, the four repeat rides BC model was able to explain only 17% (model adjusted-R^2 = 0.17) of the variance in data, while the 12 repeat rides BC model was able to explain 45% (model adjusted-R^2 = 0.45) of the variance. The highest model adjusted-R^2 was observed for the 12 repeat rides UFPs model (0.58), followed by the 12 repeat rides BC model (0.45) and the 12 repeat rides $PM_{2.5}$ model (0.29). The validation RMSE also followed the same trend. The model adjusted-R^2 and validation (10-fold CV) R^2 were very similar, indicating stable models.

On-road pollutant models included three to five final potential predictors (Table 5). In all the on-road LUR models, highway road length and tertiary road length at a small buffer of 25 m emerged as predictors. This showed that on-road air pollution levels were being influenced by spatially immediate sources. The best on-road $PM_{2.5}$ model included four predictors: highway road length in 25 m buffer, tertiary road length in 25 m buffer, NDVI in 100 m buffer, and wind speed. BC and UFPs models performed better compared to $PM_{2.5}$ models. Explanatory predictors for the best BC model were highway road length and tertiary road length in 25 m buffer and built-up area in 5000 m buffer, indicating anthropogenic sources. The best UFPs model explained 58% of the variability and included rail track length and NDVI in a 5000 m buffer along with highway road length and tertiary road length. The Moran's I statistic for model residues is also provided in Table 5.

4.5.2 Ambient $PM_{2.5}$ Models

We trained monthly, seasonal, and study period mean LUR models using the PA network ambient $PM_{2.5}$ data. For LUR training, the gridded satellite data and reanalysis data were interpolated to 50 m spatial resolution. Model performance was validated using both 10-fold cross-validation and leave-one-out cross-validation (LOOCV). LUR training and validation statistics are summarised in Table 6. Among monthly models, the highest adjusted-R^2 (0.41) was for January; the lowest adjusted-R^2 (0.15) was for May. We developed monthly models and included one to four predictors (Table 6). Variables indicative of areas sources/sinks (shrubland and bareland) emerged as final predictors in most of the monthly models. The ranked correlation between the model and measured $PM_{2.5}$ was the strongest and highest for the January model. Moderate validation R^2 (0.46) and lower RMSE (10-fold = 4.6 $\mu g\ m^{-3}$ and LOOCV = 4.9 $\mu g\ m^{-3}$) values were observed for the January model. Among seasonal models, the winter (DJF) model performed

better with higher model adjusted-R^2 (0.30) over the summer model (0.17). The performance metrics of the winter model were comparable to that of the January model, with moderate values of validation-R^2 (0.34) and lower RMSE (10-fold = 4.2 μg m^{-3} and LOOCV = 4.5 μg m^{-3}). The 6-months (DJFMAM) model performance was superior to all other models with better model adjusted-R^2 (0.41), validation-R^2 (0.45), and RMSE (10-fold = 3.9 μg m^{-3} and LOOCV = 4.2 μg m^{-3}). The Moran's I statistic for the model residues is also provided in Table 6. Moran's I values were much lower for ambient $PM_{2.5}$ models compared to on-road air pollution LUR models.

Table 5: LUR models for on-road air pollution and their descriptive statistics

Pollutant	N	Final model predictors	Adjusted-R^2 (model)	Validation-R^2 (10-CV)	RMSE (10-CV)	ρ (model)	Mean measured concentrations	Moran's I
Four repeat rides: $PM_{2.5}$ ($\mu g\ m^{-3}$)	16930	ndvi_100m + bareland_1000 + percentagewaterbody_5000 + highwayroadlength_25 + precipitation	0.15	0.16	52	0.42	89	0.18
Eight repeat rides: $PM_{2.5}$	13210	ndvi_100+ndvi_5000 + bareland_1000 + highwayroadlength_25 + tertiaryroadlength_25	0.27	0.26	30	0.61	79	0.17
12 repeat rides: $PM_{2.5}$	7837	ndvi_100 + highwayroadlength_25 + tertiaryroadlength_25 + windspeed	0.29	0.29	24	0.69	73	0.13
Four repeat rides: BC ($\mu g\ m^{-3}$)	16890	inversedistancetoindustry + highwayroadlength_25 + tertiaryroadlength_25	0.17	0.17	41	0.71	36	0.14
Eight repeat rides: BC	13205	cropland_500 + highwayroadlength_25 + tertiaryroadlength_25	0.34	0.34	27	0.75	33	0.15
12 repeat rides: BC	7836	Builtup_5000+highwayroadlength_25 + tertiaryroadlength_25	0.45	0.46	22	0.82	31	0.12
Four repeat rides: UFPs ($x\ 10^3\ cm^{-3}$)	16923	shrubland_5000 + highwayroadlength_25 + tertiaryroadlength_25	0.38	0.38	43	0.72	67	0.19
Eight repeat rides: UFPs	13176	shrubland_5000 + cropland_1000 + highwayroadlength_25 + tertiaryroadlength_25	0.47	0.47	38	0.77	68	0.17
12 repeat rides: UFPs	7780	highwayroadlength_25 + tertiaryroadlength_25 + rail_5000 + ndvi_5000	0.58	0.58	35	0.83	68	0.13

Table 6: LUR models for ambient $PM_{2.5}$ and their descriptive statistics. N represents the number of data points after removing influential observations.

Time period	N	Final model predictors	Adjusted-R^2 (model)	Validation-R^2 (10-CV)	RMSE (10-CV) (µg m^{-3})	RMSE (LOOCV) (µg m^{-3})	ρ (model)	Mean measured concentrations (µg m^{-3})	Moran's I
December 2021	43	ndvi_300 + vwind + bareland_500	0.38	0.42	4.0	4.4	0.62	45	-0.03
January 2022	54	inv_sq_dis_nearest_ind + elevation + shrubland_300 + rail_1000	0.41	0.46	4.9	4.6	0.69	39	-0.03
February 2022	45	elevation + temperature + shrubland_500 + bareland_500	0.27	0.34	4.4	4.6	0.52	43	-0.01
March 2022	45	inv_dis_airport + elevation + shrubland_100	0.25	0.31	4.3	4.8	0.61	48	0.01
April 2022	41	shrubland_500 + highwayroadlength_25	0.21	0.25	4.5	4.9	0.13	35	-0.01
May 2022	42	inv_dis_ind	0.15	0.18	4.7	5.5	0.44	26	-0.01
Winter (DJF)	44	shrubland_300 + baerland_1000	0.31	0.34	2.2	4.5	0.51	42	-0.01
Summer (MAM)	40	shrubland_5000 + bareland_500	0.17	0.22	3.8	4.2	0.47	36	-0.02
Winter + Summer (DJFMAM)	39	long + shurbland_300 + highwayroadlength_25	0.41	0.45	3.9	4.2	0.61	39	-0.02

4.6 Spatial predictions

4.6.1 On-road air pollution

The LUR models developed with the data of the 12 repeat rides were found to be best performing compared to the four and eight repeat-ride models. Hence, on-road predictions were made using the 12 repeat-ride models. Spatial predictions (~0.7 million in number) of on-road air pollution were carried out at a 50 m distance and limited to the BBMP area of the Bengaluru Urban district. The predictions represent the study period daytime mean pollutant concentrations.

The city mean–predicted on-road $PM_{2.5}$, BC, and UFPs were ~65 µg m^{-3}, 18 µg m^{-3}, and 38 × 10^3 cm^{-3}, respectively. Daytime on-road mean $PM_{2.5}$ values were higher (~65 µg m^{-3}) compared to their ambient counterparts (~45 µg m^{-3}). Figure 14 shows the spatial variation of daytime on-road $PM_{2.5}$, BC, and UFPs. Across the BBMP area, predicted $PM_{2.5}$ varied between 10 and 170 µg m^{-3}. On-road $PM_{2.5}$ exhibited higher values (between 72 and 80 µg m^{-3}) in the western parts of the BBMP area, which include Dasarahalli, Rajarajeshwari Nagar, West Zone, South Zone, and parts of Bommanahalli. As expected, $PM_{2.5}$ concentrations were higher along major roads and arterial roads. The Bengaluru–Mysore Road, NICE Road, Kanakapura Road, Magadi Main Road, and Tumkur Road were characterised by high levels of $PM_{2.5}$ (>88 µg m^{-3}). Most of the eastern parts were characterised by low levels of on-road $PM_{2.5}$ (<60 µg m^{-3}).

On-road levels of BC and UFPs followed almost a similar pattern, with lower values observed along the BBMP boundary. Higher concentrations were observed in the South and West zones. Moderate values were observed in the rest of the BBMP area. As expected, daytime on-road BC concentrations were higher (>22 µg m^{-3}) along the major and arterial road network. Higher to moderate concentrations (>15 µg m^{-3}) were observed in most of the zones (South, West, parts of Rajarajeshwari Nagar, Bommanahalli, Dasarahalli, East, and Mahadevpura). Daytime on-road UFPs were higher (>80 × 10^3 cm^{-3}) along major roads. They were between 50 × 10^3 cm^{-3} and 60 × 10^3 cm^{-3} in the West and South zones, respectively. UFPs were characterised by larger spatial variability (in terms of SD and IQR) compared to BC and $PM_{2.5}$ (see Table 7).

Table 8 summarises on-road air pollution statistics for various road types. The box plot (Figure 15) depicts the variation in $PM_{2.5}$, BC, and UFPs concentrations for different road types (major, arterial, residential, and unclassified roads). Major roads observed the highest levels of pollution ($PM_{2.5}$: ~81 µg m^{-3}, BC: ~39 µg m^{-3}, and UFPs: ~80 × 10^3 cm^{-3}), followed by arterial roads ($PM_{2.5}$: ~65 µg m^{-3}, BC: ~15 µg m^{-3}, and UFPs: ~33 × 10^3 cm^{-3}), residential roads ($PM_{2.5}$: ~60 µg m^{-3}, BC: ~12 µg m^{-3}, and UFPs: ~22 × 10^3 cm^{-3}), and unclassified roads ($PM_{2.5}$: ~58 µg m^{-3}, BC: ~13 µg m^{-3}, and UFPs: ~21 × 10^3 cm^{-3}).

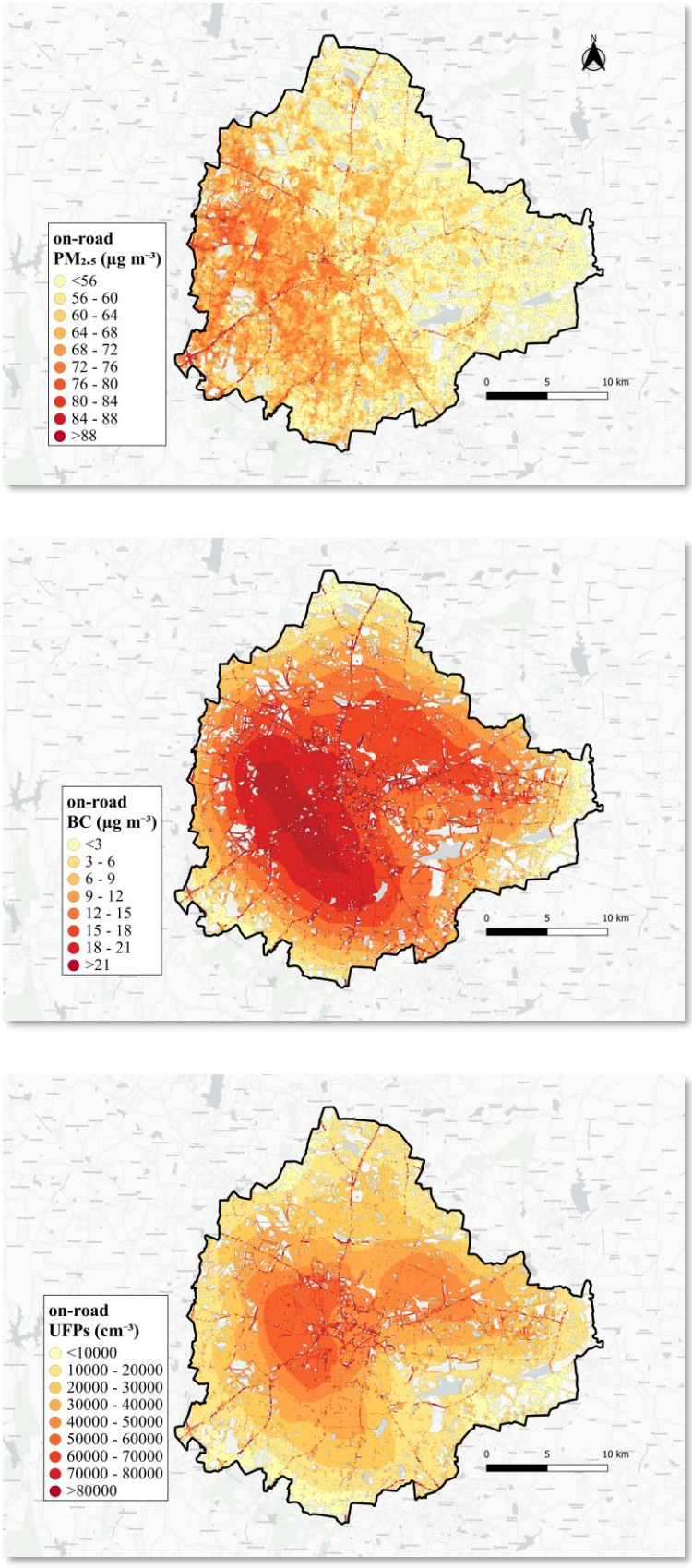


Figure 14: Spatial distribution of on-road pollution levels of $PM_{2.5}$ (top panel), BC (middle panel), and UFPs (bottom panel) as predicted by LUR models

Table 7: Predicted on-road pollution statistics

Pollutant	Mean	SD	5th percentile	25th percentile	50th percentile	75th percentile	95th percentile	Minimum	Maximum
$PM_{2.5}$ (μg m^{-3})	65	11.7	48	58	65	71	84	10	170
BC (μg m^{-3})	18	12.0	4	11	16	20	44	0.1	160
UFPs (× 10^3 cm^{-3})	38	27.0	10	20	32	48	93	0.01	313

Table 8: Predicted on-road air pollution statistics by road type

Road type	Number of predictions	Mean	SD	5th percentile	25th percentile	50th percentile	75th percentile	95th percentile	Minimum	Maximum
	$PM_{2.5}$ (μg m^{-3})									
Major	37,276	81	19	50	66	82	94	111	35	165
Arterial	155,811	65	14	43	54	65	76	87	19	160
Residential	367,253	60	11	41	52	60	68	76	23	113
Unclassified	143,661	58	15	36	46	57	69	83	16	127
	BC (μg m^{-3})									
Major	37,276	39	21	12	18	38	54	76	0.1	158
Arterial	155,811	15	13	1	4	14	23	35	0.1	149
Residential	367,253	12	7	1	6	12	17	22	0.1	84
Unclassified	143,661	13	10	1	5	13	18	27	0.1	98
	UFPs (× 10^3 cm^{-3})									
Major	37,276	80	43	11	44	82	107	157	0.01	313
Arterial	155,811	33	27	3	13	24	47	81	0.01	296
Residential	367,253	22	16	2	90	19	33	53	0.01	165
Unclassified	143,661	21	22	0.8	47	13	34	57	0.01	217

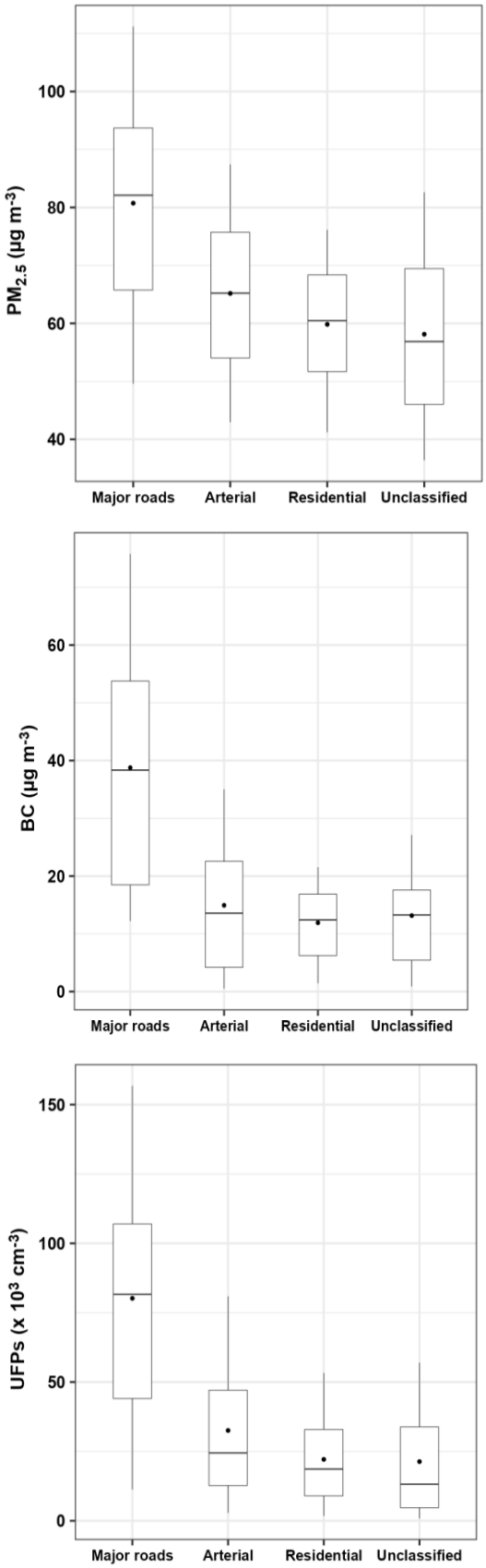


Figure 15: Distribution of predicted on-road pollutant levels by road type. The central line of the box represents the median, the dot represents the mean, the box represents the inter-quartile range (IQR), and the whiskers represent the 5th and 95th percentiles. Outliers are omitted for readability.

4.6.2 Ambient $PM_{2.5}$

Ambient LUR model predictions were used to map $PM_{2.5}$ at a very high spatial resolution over the Bengaluru Urban and Rural districts. Spatial predictions were carried out at locations 50 m apart (~1.8 million points covering the study region). We predicted ambient $PM_{2.5}$ using models characterised with a minimum of 0.30 model adjusted-R^2 value. Following the criteria, we were left with four models pertaining to December 2021, January 2022, winter (DJF), and winter and summer combined (DJFMAM; see Figures 16 and 17). We also removed possible outliers from predictions, following the method suggested by Henderson et al. (2007). A lower and upper limit equal to the lowest −50% and the highest +50% $PM_{2.5}$ values from the monitoring campaign was set, and any values beyond these thresholds were removed from the distribution.

Across the study months and seasons, predicted $PM_{2.5}$ values ranged between 10 and 100 μg m^{-3} within the study region. The mean $PM_{2.5}$ was comparable across the Bengaluru Rural district (ranging between 33 μg m^{-3} and 42 μg m^{-3}) and the Bengaluru Urban district (35 μg m^{-3} to 44 μg m^{-3}). Within the BBMP region, values ranged between 38 μg m^{-3} and 45 μg m^{-3}. We observed very small differences in the predicted $PM_{2.5}$ between Bengaluru Urban and Rural districts. Detailed statistics are provided in Table 9. Differences in the study period mean and median $PM_{2.5}$ values over Bengaluru Rural, Bengaluru Urban, and BBMP regions were less than the LUR model RMSE values.

The mean $PM_{2.5}$ was the highest in December 2021, followed by winter 2021–2022, January 2022, and DJFMAM. In December 2021, high $PM_{2.5}$ concentrations (>50 μg m^{-3}) were observed in the eastern part of the study region. Isolated hotspots in the western peri-urban region coincided with stone-crushing locations. The January 2022 map indicated higher concentrations within the Bengaluru Urban district and also in the BBMP region. Parts of Mahadevpura, Dasarahalli, Rajarajeshwari Nagar, West Zone, and South Zone showed clusters of higher $PM_{2.5}$ during January 2022. Bommasandra and Attibele areas were also characterised by higher $PM_{2.5}$ (>50 μg m^{-3}). Lower $PM_{2.5}$ (~33 μg m^{-3}) was observed in the Bengaluru Rural district (except a hotspot region in the northern part of the district).

During winter, a few rural and peri-urban regions were characterised by higher $PM_{2.5}$ (Figure 17). Peri-urban hotspots coincided with stone-crushing locations (similar to that observed in December 2021). Winter spatial mean $PM_{2.5}$ values for the Bengaluru Urban district (39 μg m^{-3}), BBMP region (41 μg m^{-3}), and Bengaluru Rural district (38 μg m^{-3}) were comparable. The study period mean spatial map (winter and summer seasons combined, DJFMAM) exhibited a gradient in $PM_{2.5}$, which increased towards the eastern parts of the study region.

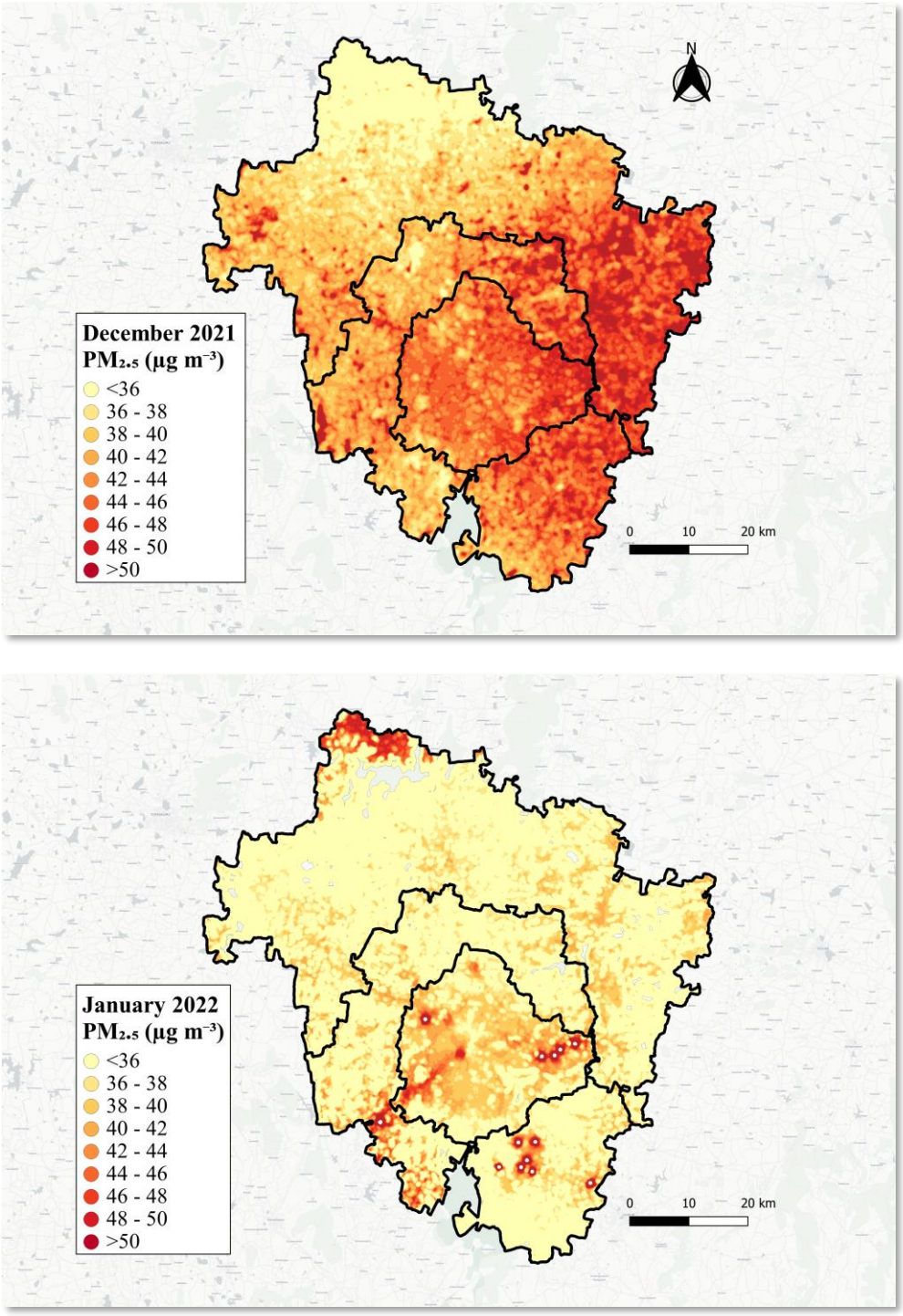


Figure 16: Spatial distribution of predicted ambient $PM_{2.5}$ for the months of December 2021 (top panel) and January 2022 (bottom panel)

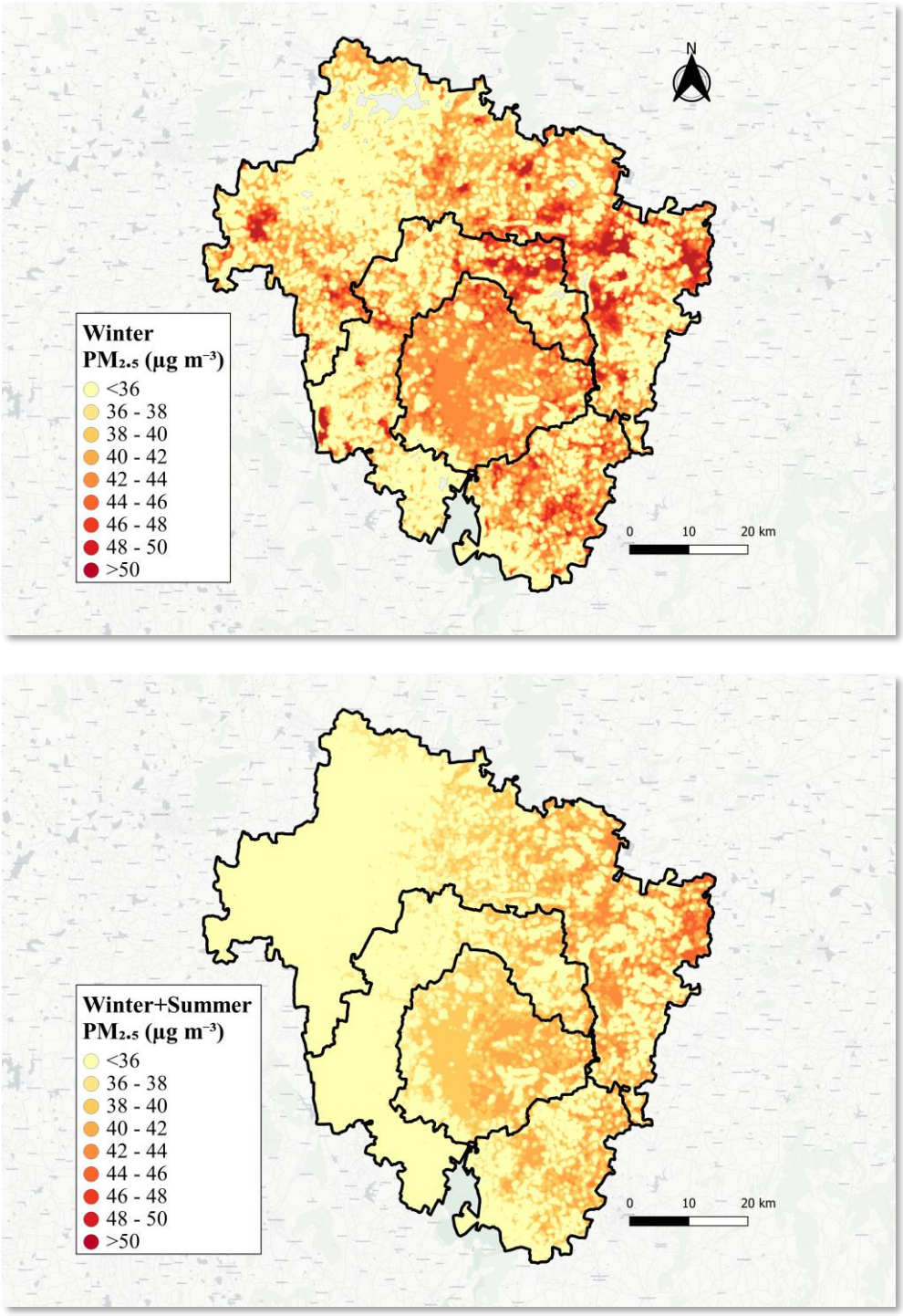


Figure 17: Spatial distribution of predicted ambient $PM_{2.5}$ for the winter period (DJF; top panel) and winter plus summer period (DJFMAM; bottom panel)

Table 9: Predicted ambient $PM_{2.5}$ statistics by region

Month/season	Mean (µg m^{-3})	SD (µg m^{-3})	5th percentile (µg m^{-3})	25th percentile (µg m^{-3})	50th percentile (µg m^{-3})	75th percentile (µg m^{-3})	95th percentile (µg m^{-3})	Minimum (µg m^{-3})	Maximum (µg m^{-3})
Bengaluru Urban district									
December 2021	44	3.9	39	47	44	47	50	34	77
January 2022	35	7.3	21	32	37	40	47	11	83
Winter (DJF) 2021–2022	39	7.1	25	36	41	44	47	11	72
DJFMAM 2021–2022	35	6.8	24	32	37	40	42	9	92
Bruhat Bengaluru Mahanagara Palike (BBMP)									
December 2021	45	2.7	40	44	45	47	50	36	64
January 2022	38	6.1	27	36	39	41	46	11	83
Winter (DJF) 2021–2022	41	5.1	30	41	43	44	46	11	52
DJFMAM 2021–2022	38	5.8	30	37	39	40	42	13	92
Bengaluru Rural district									
December 2021	42	5.7	34	38	41	46	52	29	74
January 2022	33	7.2	18	29	34	38	41	11	52

Month/season	Mean (µg m^{-3})	SD (µg m^{-3})	5th percentile (µg m^{-3})	25th percentile (µg m^{-3})	50th percentile (µg m^{-3})	75th percentile (µg m^{-3})	95th percentile (µg m^{-3})	Minimum (µg m^{-3})	Maximum (µg m^{-3})
Winter (DJF) 2021–2022	38	8.2	21	34	40	43	48	11	65
DJFMAM 2021–2022	34	8.2	19	30	35	39	43	9	92
Bengaluru peri-urban (region between BBMP and Bengaluru Rural)									
December 2021	44	3.9	39	42	44	47	50	34	77
January 2022	35	7.3	21	32	37	40	45	11	83
Winter (DJF) 2021–2022	39	7.1	25	36	41	44	47	11	72
DJFMAM 2021–2022	35	6.8	24	32	37	40	42	9	92
Whole study area (Bengaluru Urban + Bengaluru Rural districts)									
December 2021	43	5.0	35	39	43	46	51	29	77
January 2022	34	7.4	19	30	36	39	43	11	83
Winter (DJF) 2021–2022	39	7.7	22	35	40	43	48	11	72
DJFMAM 2021–2022	35	7.6	21	31	36	39	43	9	92

5. Discussion

The current study is one of the most comprehensive studies in terms of methodologies used, area covered, and pollutants measured in an LMIC urban setting. We sampled three air pollutants over ~1000 km (~10% of the city roads) in Bengaluru, covering different road types and neighbourhoods. The mobile monitoring was complemented by a city-wide 55-node stationary low-cost sensor network for ambient $PM_{2.5}$ measurements that covered different land-use types.

Ultra-high-resolution mapping of Bengaluru at this scale has never been done before. Pollution maps of all pollutants follow a predictable pattern, with the highest concentrations on highways and main roads. However, clear gradients appear as we move away from main roads. There is a large spatial variability within neighbourhoods. This is in agreement with findings from Apte et al. (2017) in Oakland, where pollutant concentrations varied up to eight-fold times within the same city block. Spatial representativeness of the maps increases with increasing number of rides as maps become more stable.

Differences in on-road concentrations by road type were more pronounced in UFPs and BC and subtle for $PM_{2.5}$. This could be because $PM_{2.5}$ is a regional pollutant with natural and anthropogenic sources, while BC and UFPs are more local. UFPs and BC prediction maps correlated highly and revealed higher concentrations in the western part of the city. Additionally, prediction maps of $PM_{2.5}$ across all months and seasons reveal that both urban and rural regions are equally polluted.

In all on-road LUR models, road features such as highway road length and tertiary road length within 25 m were found to be significant predictors. This shows that on-road air pollution levels are being influenced by spatially immediate sources. Most models also included indicators for green space or shrubland as significant predictors. UFPs models were the best performers and could explain 58% variability. Results indicate that BC and UFPs models are superior to $PM_{2.5}$ models, reflecting the limited predictability of the latter. Pollutants with higher spatial variability are often better predicted by LUR models. As seen in the prediction maps, $PM_{2.5}$ is more spatially homogeneous when compared to BC and UFPs maps. This phenomenon can also be observed in 'data only' maps. $PM_{2.5}$ maps show the least spatial variability, followed by UFPs and BC. The results are in agreement with Blanco et al.'s (2022) mobile monitoring work from Seattle, which shows the highest spatial variability in UFPs (17%), followed by BC (8%) and $PM_{2.5}$ (2%).

We found that LUR models for all pollutants improved (in terms of adjusted-R^2 and RMSE) with increasing number of rides. We covered the entire mobile monitoring route at least four times. To identify the optimal number of passes required for robust estimates, we covered a subset of the roads eight and 12 times. Progressively, fewer routes were sampled for eight and 12 ride passes to optimise resources. In other words, we found that capturing more temporal variation (increasing the number of passes on selected roads) rather than spatial variation (driving on more roads for fewer passes) leads to better performance of LURs. In our study, LUR models that were trained on data from 12 rides (with fewer roads covered) resulted in better models than those trained on data from four rides (with more roads covered). These findings are consistent with that of Messier et al. (2018) in Oakland where models were more sensitive to number of passes than road coverage.

Average $PM_{2.5}$ values for the stationary network across all 55 sites ranged between 36 μg m^{-3} and 42 μg m^{-3} for winter and summer, respectively. $PM_{2.5}$ levels and their spatial variability were relatively higher during the winter months compared to the summer months. This could be

because of increased anthropogenic activities, low temperatures, and stagnant weather conditions during winters. In comparison, on-road $PM_{2.5}$ values were higher than those from ambient stationary sensor networks. This could be because of the following reasons: (i) the difference in the extent of the study region (on-road measurements were limited to urban areas, while ambient measurements were carried out in peri-urban and rural areas of Bengaluru), (ii) on-road measurements were limited to daytime, while ambient measurements were performed round the clock, or (iii) the source proximity was much closer to the instrument inlet compared to that of the ambient $PM_{2.5}$ during mobile monitoring. A list of advantages and disadvantages and a comprehensive literature review of mobile monitoring LUR models are given in Xu et al. (2021).

LUR models from the stationary network data reveal that green spaces (shrubland) and road features are significant predictors of $PM_{2.5}$. Winter models performed better than summer models and explained ~40% of spatial variability in $PM_{2.5}$. Being able to build an LUR model using data from the LCS network is one of the possible practical applications of such networks to understand the spatial distribution of pollution levels. Surprisingly, prediction maps of $PM_{2.5}$ across all months and seasons reveal that both urban and rural regions are equally polluted.

However, overall, $PM_{2.5}$ models from both stationary and mobile measurements have limited predictability (lower adjusted-R^2) compared to BC and UFPs models. This could be because of (i) less variability in spatial $PM_{2.5}$ measurements, (ii) heterogeneous pollution sources that are not explained by predictors, (iii) a lack of background monitoring stations to account for the transported component of $PM_{2.5}$, (iv) rapidly varying (in space and time) $PM_{2.5}$ sources and their strengths, or (v) a lack of data on relevant predictor variables (e.g., traffic data). In areas with homogeneous sources of $PM_{2.5}$, LUR models show better predictive capability. For example, in the Cardiovascular Health Effects of Air pollution in Telangana, India (CHAI), project from rural south India, both $PM_{2.5}$ and BC LUR models performed much better and were able to explain ~58% and 79% variability in $PM_{2.5}$ and BC, respectively, (Sanchez et al., 2018) with biomass burning as the primary source of particulate matter. Accounting for diurnal patterns may lead to the training of better-performing models, especially for $PM_{2.5}$. For example, in Delhi, Saraswat et al. (2013) accounted for strong temporal (diurnal) patterns by developing different LUR models for morning and afternoon with better success (R^2 >0.7 for the $PM_{2.5}$ model). In the current study, we used hourly factors to correct the mobile monitoring data to daytime mean values.

A major challenge in LUR modelling is the availability of reliable predictor data. In the ESCAPE project, researchers used fewer sites (~20) for each area and developed LUR models of 20 different European study areas. R^2 for $PM_{2.5}$ LUR models ranged between 0.35 and 0.94, with a median of 0.71. Models with lower R^2 either had minimal variation in measurements or had limited availability of predictor data, especially traffic intensity. In current models, we use road type and length as proxies for traffic data, but it is possible that variations in traffic intensity may not be completely captured. Availability of data on traffic intensity, vehicle type, or density may help characterise this important source further and improve the prediction capability of models. Additionally, it was hard to find a background site near our study area to capture the levels of the transported component of $PM_{2.5}$.

This report showcases results from data collected until May 2022 for the performance evaluation of LCSs. The comparison of LCS performances shows little seasonality between winter and summer months. As the next step and continuation of project activities, we will continue to collect data for the rest of the year, which will allow us to compare sensor performance across all seasons. Similarly, we are planning to continue collecting network ambient $PM_{2.5}$ data and build LUR

models for the rest of the year, which will allow us to predict seasonal and annual ambient $PM_{2.5}$. This study is limited to particulate matter monitoring using ambient and mobile monitoring approaches. Future studies may focus on the characterisation of gaseous criteria pollutants by employing multi-pollutant sensors in ambient sites. Measuring multiple pollutants together can be a cost-effective way to help inform more comprehensive pollution mitigation policies.

5.1 Policy implications

Based on the current study findings, the following policy recommendations are proposed:

1. Prioritise mitigation activities abating vehicular emissions in the western parts of the BBMP area.
2. Adopt regional strategies for the reduction of air pollution rather than opting for city-specific action plans. ($PM_{2.5}$ is a regional pollutant, and only minute differences were observed in spatial $PM_{2.5}$ levels across urban, peri-urban, and rural areas of Bengaluru.)
3. Establish representative regulatory air pollution monitoring stations in rural and peri-urban areas.
4. Audit unorganised polluting industries/activities in the outskirts of Bengaluru.

We accomplished high-resolution city-wide air pollution mapping through a combination of direct measurements and statistical modelling. An integrated monitoring framework that involves strategically placed LCSs and occasional high-resolution mapping can inform and complement the evolving traditional reference-grade monitoring systems. With a dense network of PurpleAir sensors, we had 1 $PM_{2.5}$ sensor per 17.6 km^2 compared to 1 monitor per 92.6 km^2 (the density of reference-grade monitors in the city).

High-resolution city-wide pollution maps from mobile monitoring and LUR model predictions reveal spatial gradients and pollution hotspots. Identifying these hotspots can be the first step toward geographically targeted pollution mitigation efforts. These pollution maps can also be overlaid with locations of schools, hospitals, and nursing homes to characterise the air pollution exposure of vulnerable populations to inform pollution mitigation policies. Land use varies across the city, and non-residential on-road sampling reflects exposure of specific populations—such as roadside store owners, street vendors, and traffic police—to pollutants apart from short-term commuters. People involved in driving-related occupations such as e-commerce delivery services, taxi services, and auto drivers are also exposed to pollutants.

Since these methods can capture spatio-temporal variations, they can be used in several policy-relevant applications. While on-road measurements allow for the characterisation of air pollution exposure by measuring levels closer to where people live and breathe, ambient measurements can be useful for regulatory purposes. For example, characterising spatial patterns of $PM_{2.5}$ levels using LCSs can inform the siting for installing reference-grade sensors in cities with no regulatory monitoring. For cities with some level of regulatory monitoring, it can help inform the siting of new reference sensors to fill data gaps. Given the portability and ease of installation and operability of low-cost sensors, they can also potentially help in evaluating geographically targeted pollution mitigation measures by measuring pollution levels before, during, and after the implementation of mitigation measures. Identifying pollution hotspots could help locate specific sources to prioritise local action and resources. Exposure data collected from these methods can be used for health impact assessments to estimate the health and economic burden of air pollution on a particular city or population, thereby helping policymakers make informed decisions while allocating scarce resources.


PM2.5

6. Limitations and Challenges

6.1 Mobile Monitoring

Mobile monitoring in a highly populated urban centre such as Bengaluru poses unique challenges for project implementation. First, it can be resource intensive. Wall-to-wall driving, where we plan to cover every lane to capture neighbourhood-level variability, sometimes requires navigating through tightly packed lanes with vehicles parked on both sides. At times, we have had to skip roads completely that look viable on the map but are not practically navigable because of limited space. Also, factors such as unplanned and indefinite road closures and poor road conditions led to occasional changes in the study routes. Such uncertainties call for on-the-fly decision-making, necessitating a research field staff to accompany the driver during all mobile monitoring rides. The field staff also helps keep a check on instrument functioning to ensure data quality.

Second, air-pollution measurements, especially from optical instruments, on a mobile platform are susceptible to vibration-related noise because of (poor) road conditions and vehicle suspension effectiveness. In our study, we applied appropriate corrections for the BC measurement to remove vibration-related artefacts.

Third, our monitoring platform (car) was fuelled by CNG, which is low-emitting but not completely emissions-free. Emissions from the vehicle can bias the measurements to some extent. An electric vehicle (i.e., a zero-tailpipe emissions vehicle) would be the best choice for a mobile platform, but the battery capacity of the vehicle limits the number of monitoring kilometres per session.

Finally, there are practical constraints on the range of data collection. Bengaluru has a long rainy season, limiting the number of monitoring days. Excluding rainy days may also lead to an overestimation of average pollution concentrations. Because of safety reasons, mobile monitoring was conducted only during daytime. Thus, night-time concentrations are not captured in this campaign. High traffic and poor road conditions also lead to low average driving speeds, limiting the distance covered.

6.2 Stationary network

Most of the PA sensors in the project were hosted in residential buildings and institutions such as college and school campuses. Regular maintenance and upkeep of sensors depend on getting access to the site premises promptly. However, because of Covid-19 (coronavirus disease) restrictions, some of the sites could not be accessed for extended durations. Most schools and colleges moved to the online mode of instruction because of safety concerns and lockdowns; therefore, we did not always have access to sensor sites. Apart from pandemic-related constraints, a host's enthusiasm and attitude toward the project may also change because of various reasons that are outside our control. For long projects, a sense of fatigue may set in, impacting the host's relationship with the project team. We were not able to access sensor sites in such cases. Despite these challenges, we were able to retrieve more than 75% of sensor data.

Given that aerosol properties change with time and space, it is ideal to perform LCS $PM_{2.5}$ calibration over a range of concentrations and conditions that captures variations in the composition of PM within the study region. However, because of resource constraints, we trained the LCS calibration model using just one collocation site. Additionally, depending on the availability and location of host sites, some monitors were installed above the prescribed height of 10 m above the ground.

Lastly, a common limitation for both mobile and stationary monitoring projects is the absence of an emission-free background reference site. This can impact the LUR modelling in precisely accounting for the transported component or pollution.

7. Summary

The current study uses an LCS network, regulatory data, and mobile monitoring data to develop high-resolution spatial air pollution maps concurrently in the same urban geography. The mobile monitoring campaign included repeat measurements of on-road $PM_{2.5}$, BC, and UFPs over ~10% of Bengaluru roads, covering all major neighbourhoods and road types. More than 2 million data points were collected during the study. Quality-checked data were gridded onto 50 m road segments and used to build LUR models. We also established a city-wide 55-node PurpleAir LCS network covering urban, peri-urban, and rural parts of Bengaluru. The network measurements of ambient $PM_{2.5}$ were carried out during the winter and summer seasons of 2021–2022. We also collocated various LCSs with BAM to investigate the accuracy of the $PM_{2.5}$ measurements. The key results obtained from the study are summarised below.

- Spatial variability in $PM_{2.5}$ observations was higher during winter compared to summer. The seasonal mean $PM_{2.5}$ was around 42 µg m^{-3} and 36 µg m^{-3} for winter and summer, respectively. Monthly and seasonally aggregated $PM_{2.5}$ values were used to build LUR models.
- Across the study period, predicted ambient $PM_{2.5}$ values ranged between 10 and 100 µg m^{-3}. Most of the study region was characterised by $PM_{2.5}$ levels less than the daily national standard (60 µg m^{-3}), and around half of the spatial $PM_{2.5}$ values were less than the annual national standard (40 µg m^{-3}).
- On-road air pollution levels were found to be 2–10 times higher than ambient pollution levels. For the study routes, the average $PM_{2.5}$, BC, and UFPs concentrations were found to be 73 µg m^{-3}, 31 µg m^{-3}, and 68 x 10^3 cm^{-3}, respectively. The representativeness of on-road pollution levels increased with the increasing number of repeat rides. The highest on-road pollution levels were observed over major roads, followed by arterial and residential roads, and the gradient was shallower for $PM_{2.5}$ compared to other pollutants.
- For LUR models developed using mobile monitoring air pollution data, area sources/sinks emerged as potential predictors for most of the ambient $PM_{2.5}$ models, while areas and line sources/sinks were observed as final predictors for the mobile monitoring air pollution models.
- LUR models trained on mobile monitoring data showed better performance compared to ambient LUR models in terms of the model adjusted-R^2. In terms of RMSE, ambient models performed better. Among ambient $PM_{2.5}$ models, winter models performed better than summer models. Among on-road air pollution models, BC and UFPs models performed better than $PM_{2.5}$ models. Models trained using air pollution data from 12 repeat rides were able to explain the variance in the data better.
- The daytime on-road mean $PM_{2.5}$, BC, and UFPs values were ~65 µg m^{-3}, 18 µg m^{-3}, and 38 × 10^3 cm^{-3}, respectively. Major roads (including highways and the Outer Ring Road) were characterised by higher levels of pollution ($PM_{2.5}$: ~81 µg m^{-3}, BC: ~39 µg m^{-3}, and UFPs: ~80 × 10^3 cm^{-3}) compared to arterial, residential, and unclassified roads. The difference in concentrations by road type is more pronounced for BC and UFPs than for $PM_{2.5}$.
- Based on the collocation (BAM and LCSs) data analysis, we observed that a multilinear model (with RH and $PM_{2.5}$ as predictors) performed best and was further used for correcting ambient PA $PM_{2.5}$ measurements. The RMSE in the uncorrected PA $PM_{2.5}$ was reduced by almost three times after applying the correction model.

- Ten different PM LCS were evaluated, and the bias (RMSE) in the hourly $PM_{2.5}$ measurements ranged between 8 and 34 μg m^{-3}. The bias was marginally reduced when daily means were compared instead of hourly means. Plantower-based LCSs performed relatively better. Not much seasonality was observed in the performance of LCSs.

8. References

Apte, J. S., Kirchstetter, T. W., Reich, A. H., Deshpande, S. J., Kaushik, G., Chel, A., ... & Nazaroff, W. W. (2011). Concentrations of fine, ultrafine, and black carbon particles in auto-rickshaws in New Delhi, India. *Atmospheric Environment*, *45*(26), 4470-4480. https://doi.org/10.1016/j.atmosenv.2011.05.028

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., ... & Hamburg, S. P. (2017). High-resolution air pollution mapping with Google street view cars: exploiting big data. *Environmental Science & Technology*, *51*(12), 6999-7008. https://doi.org/10.1021/acs.est.7b00891

Ban-Weiss, G. A., Lunden, M. M., Kirchstetter, T. W., & Harley, R. A. (2009). Measurement of black carbon and particle number emission factors from individual heavy-duty trucks. *Environmental Science & Technology*, *43*(5), 1419-1424. https://doi.org/10.1021/es8021039

Barkjohn, K. K., Gantt, B., & Clements, A. L. (2021). Development and application of a United States-wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor. *Atmospheric Measurement Techniques*, *14*(6), 4617-4637. https://doi.org/10.5194/amt-14-4617-2021

Blanco, M. N., Gassett, A., Gould, T., Doubleday, A., Slager, D. L., Austin, E., ... & Sheppard, L. (2022). Characterization of Annual Average Traffic-Related Air Pollution Concentrations in the Greater Seattle Area from a Year-Long Mobile Monitoring Campaign. *Environmental Science & Technology*. https://doi.org/10.1021/acs.est.2c01077

Bové, H., Bongaerts, E., Slenders, E., Bijnens, E. M., Saenen, N. D., Gyselaers, W., ... & Nawrot, T. S. (2019). Ambient black carbon particles reach the fetal side of human placenta. *Nature communications*, *10*(1), 1-7. https://doi.org/10.1038/s41467-019-11654-3

Brauer, M., Guttikunda, S. K., Nishad, K. A., Dey, S., Tripathi, S. N., Weagle, C., & Martin, R. V. (2019). Examination of monitoring approaches for ambient air pollution: A case study for India. *Atmospheric Environment*, *216*, 116940. https://doi.org/10.1016/j.atmosenv.2019.116940

CSTEP & ILK Labs (2020). Mobile-monitoring campaign for air-pollution studies in Bengaluru. CSTEP-WS-2020-04.

CSTEP & ILK Labs (2022a). Best practices for deploying and maintaining a low-cost $PM_{2.5}$ sensor network. CSTEP-WS-2022-02.

CSTEP & ILK Labs (2022b). Performance assessment of low-cost $PM_{2.5}$ sensors. CSTEP-WP-2022-01.

Duvall, R., Clements, A., Hagler, G., Kamal, A., Kilaru, V., Goodman, L., ... & Dye, T. (2021). Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors: Use in Ambient, Outdoor, Fixed Sites, Non-Regulatory Supplemental and Informational Monitoring Applications. *US EPA Office of Research and Development*. https://cfpub.epa.gov/

Eeftens, M., Beelen, R., De Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., ... & Hoek, G. (2012). Development of land use regression models for $PM_{2.5}$, $PM_{2.5\ absorbance}$, PM_{10} and PM_{coarse} in 20 European study areas; results of the ESCAPE project. *Environmental Science & Technology*, *46*(20), 11195-11205. https://doi.org/10.1021/es301948k

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... & Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, *45*(2). https://doi.org/10.1029/2005RG000183

Ganguly, T., Selvaraj, K. L., & Guttikunda, S. K. (2020). National Clean Air Programme (NCAP) for Indian cities: review and outlook of clean air action plans. *Atmospheric Environment: X*, *8*, 100096. https://doi.org/10.1016/j.aeaoa.2020.100096

Gascon, M., Cirach, M., Martínez, D., Dadvand, P., Valentín, A., Plasència, A., & Nieuwenhuijsen, M. J. (2016). Normalized difference vegetation index (NDVI) as a marker of surrounding greenness in epidemiological studies: The case of Barcelona city. *Urban Forestry & Urban Greening*, *19*, 88-94. https://doi.org/10.1016/j.ufug.2016.07.001

Hoek, G., Beelen, R., De Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, *42*(33), 7561-7578. https://doi.org/10.1016/j.atmosenv.2008.05.057

Janssen, N. A., Hoek, G., Simic-Lawson, M., Fischer, P., Van Bree, L., Ten Brink, H., ... & Cassee, F. R. (2011). Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM_{10} and $PM_{2.5}$. *Environmental health perspectives*, *119*(12), 1691-1699. https://doi.org/10.1289/ehp.1003369

L'Orange, C., Neymark, G., Carter, E., & Volckens, J. (2021). A High-throughput, Robotic System for Analysis of Aerosol Sampling Filters. *Aerosol and Air Quality Research*, *21*, 210037. https://doi.org/10.4209/aaqr.210037

Messier, K. P., Chambliss, S. E., Gani, S., Alvarez, R., Brauer, M., Choi, J. J., ... & Apte, J. S. (2018). Mapping air pollution with Google Street View cars: Efficient approaches with mobile monitoring and land use regression. *Environmental Science & Technology*, *52*(21), 12563-12572. https://doi.org/10.1021/acs.est.8b03395

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., ... & Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349-4383. https://doi.org/10.5194/essd-13-4349-2021

Ohlwein, S., Kappeler, R., Kutlar Joss, M., Künzli, N., & Hoffmann, B. (2019). Health effects of ultrafine particles: a systematic literature review update of epidemiological evidence. *International journal of public health*, *64*(4), 547-559. https://doi.org/10.1007/s00038-019-01202-7

Pandey, A., Brauer, M., Cropper, M. L., Balakrishnan, K., Mathur, P., Dey, S., ... & Dandona, L. (2021). Health and economic impact of air pollution in the states of India: the Global Burden of Disease Study 2019. *The Lancet Planetary Health*, *5*(1), e25-e38. https://doi.org/10.1016/S2542-5196(20)30298-9

Rivas, I., Mazaheri, M., Viana, M., Moreno, T., Clifford, S., He, C., ... & Querol, X. (2017). Identification of technical problems affecting performance of DustTrak DRX aerosol monitors. *Science of the Total Environment*, *584*, 849-855. https://doi.org/10.1016/j.scitotenv.2017.01.129

Sanchez, M., Ambros, A., Milà, C., Salmon, M., Balakrishnan, K., Sambandam, S., ... & Tonne, C. (2018). Development of land-use regression models for fine particles and black carbon in peri-urban South India. *Science of the Total Environment*, *634*, 77-86. https://doi.org/10.1016/j.scitotenv.2018.03.308

Saraswat, A., Apte, J. S., Kandlikar, M., Brauer, M., Henderson, S. B., & Marshall, J. D. (2013). Spatiotemporal land use regression models of fine, ultrafine, and black carbon particulate matter in New Delhi, India. Environmental Science & Technology, 47(22), 12903-12911. https://doi.org/10.1021/es401489h

Schraufnagel, D. E. (2020). The health effects of ultrafine particles. *Experimental & Molecular Medicine*, *52*(3), 311-317. https://doi.org/10.1038/s12276-020-0403-3

Shiva Nagendra, S. M., Yasa, P. R., Narayana, M. V., Khadirnaikar, S., & Rani, P. (2019). Mobile monitoring of air pollution using low cost sensors to visualize spatio-temporal variation of pollutants at urban hotspots. *Sustainable Cities and Society*, *44*, 520-535. https://doi.org/10.1016/j.scs.2018.10.006

Van Geffen, J., Boersma, K. F., Eskes, H., Sneep, M., Ter Linden, M., Zara, M., & Veefkind, J. P. (2020). S5P TROPOMI NO_2 slant column retrieval: Method, stability, uncertainties and comparisons with OMI. *Atmospheric Measurement Techniques*, *13*(3), 1315-1335. https://doi.org/10.5194/amt-13-1315-2020

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., ... & Levelt, P. F. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, *120*, 70-83. https://doi.org/10.1016/j.rse.2011.09.027

Viana, M., Rivas, I., Reche, C., Fonseca, A. S., Pérez, N., Querol, X., ... & Sunyer, J. (2015). Field comparison of portable and stationary instruments for outdoor urban air exposure assessments. *Atmospheric Environment*, *123*, 220-228. https://doi.org/10.1016/j.atmosenv.2015.10.076

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., ... & Arino, O. (2021). ESA WorldCover 10 m 2020 v100. *Zenodo.* https://doi.org/10.5281/zenodo.5571936

9. Appendix A

Table A1: Mobile monitoring route details

Route name	Number of repeat rides	Actual road length (km)	Unique road length (km)	Major roads (km)	Arterial roads (km)	Residential roads (km)	Unclassified roads (km)	Areas covered
MAL	12	90	79	27	28	18	7	Malleshwaram, Vayalikaval, Sadashiv Nagar, New BEL Road, Central Business Road
KAN	8	140	102	24	68	5	5	Kannuru, Thanisandra Main Road, Hennur Main Road, Ramamurthy Nagar, Old Madras Road, Bellary Road
IND	12	105	74	17	27	29	1	Indiranagar, Marathahalli, ITPL Main Road, Whitefield Main Road
HSR	4	140	88	33	29	25	2	HSR Layout, Sarjapur Road, SH-35, Whitefield
PNY	4	90	54	4	25	9	17	Peenya Industrial Area

RRN	12	130	117	27	27	40	22	RR Nagar, Mysore Road, Turahalli Forest, Kanakpura Road, Bannerghatta Road
YLK	12	90	63	4	27	15	18	Yelahanka New Town, Judicial Layout, SH-9
ORR	12	105	88	56	10	20	2	Outer Ring Road, Dollars Colony, Kammanahalli
VIJ	8	122	89	60	16	12	1	Vijayanagar, NICE Road, Hosur Road, Chord Road, HMT Main Road
JAY	8	120	74	14	33	24	2	Jayanagar, Lalbagh Main Road, Marigowda Road
CRN	4	95	49	8	23	12	6	C V Raman Nagar, Vignan Nagar, Suranjan Das Road, Kasturi Nagar Main Road
Total	96	1227	877	274	313	209	83	

*Major roads = primary + highways+ primary trunk + motor ways (OSM classification); arterial = secondary + tertiary + motor link + primary link + second link + trunk link + service; and residential roads = residential + living street

Table A2: Technical specifications of LCSs under evaluation

Sensor	Model	Display	Rated voltage (V)	Rated current (mA)	Battery	Local storage	GPS module	Size (mm)
BlueSky	TSI-8143	No	5	100	No	Yes	No	160 x 140 x 120
Airveda	PM2510THWP-HC	No	220	NA	Yes	Yes	No	240 x 160 x 90
Aerogram	Eziostat	No	NA	NA	Optional	Yes	No	115 x 65 x 25
Prkruti	-	No	10	500	Yes	Yes	Yes	450 x 450 x 200
Atmos -I	Plantower-based	Yes	5 V DC	1500 max	Yes	Yes	Yes	420 x 200 x 140
Atmos II	Sensirion-based	Yes	5 V DC	1500 max	Yes	Yes	Yes	420 x 200 x 140
Prana Air	Outdoor Air Quality Monitor	No	5	70	Yes	Yes	Optional	150 x 50 x 200
PurpleAir	PA-II-SD	No	5	180	No	Yes	No	90 x 90 x 125
PAQS	Indoor monitor	Yes	NA	NA	Yes	Yes	No	220 x 180 x 50
Aurassure	-	Yes	12–24	1000	Optional	Optional	No	110 x 100 x 40

Table A3: Operating specifications of LCSs under evaluation

Sensor	Laser counter	Pollutant	$PM_{2.5}$ measurement range ($\mu g/m^3$)	Logging interval (minutes)	Meteorological sensor	Temperature measurement range (°C)	RH measurement range (%)	Remarks
BlueSky	Sensirion	PM_1, $PM_{2.5}$, PM_4, PM_{10}	0–1000	1	Yes	-40 to 125	0 to 95	Additional subscription for high-resolution data
Airveda	Nova	$PM_{2.5}$, PM_{10}	0–999	30	Yes	10 to 60	NA	Heated inlet, additional subscription for high-resolution data
Aerogram	Plantower	PM_1, $PM_{2.5}$, PM_{10}	0–1000	0.5	Yes	-40 to 85	NA	-
Prkruti	Winsen	$PM_{2.5}$, PM_{10}	0–500	15	Yes	-40 to +125	0 to 100	-
Atmos I	Plantower	$PM_{2.5}$, PM_{10}	<1000	1	Yes	-10 to 60	0 to 99	-
Atmos II	Sensirion	$PM_{2.5}$, PM_{10}	<1000	1	Yes	-10 to 60	0 to 99	-
Prana Air	PAS-OUT-01	PM_1, $PM_{2.5}$, PM_{10}	0-1500	0.5	Yes	0 to 60	5 to 95	-
PurpleAir	Plantower	PM_1, $PM_{2.5}$, PM_{10}	0–1000	2	Yes	-40 to 85	0 to 100	-
PAQS	Honeywell	$PM_{2.5}$, PM_{10}	0–900	30	Yes	-10 to 100	0 to100	-
Aurassure	Plantower	$PM_{2.5}$, PM_{10}	<1000	1	Yes	0 to 60	<85%	-

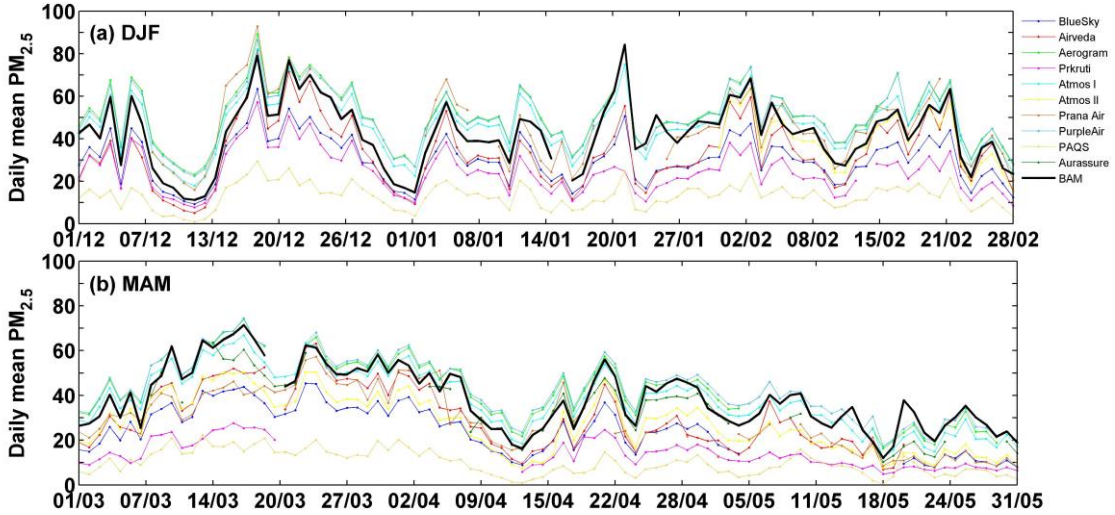


Figure A1: Time series of daily mean ambient $PM_{2.5}$ from collocated LCSs and BAM

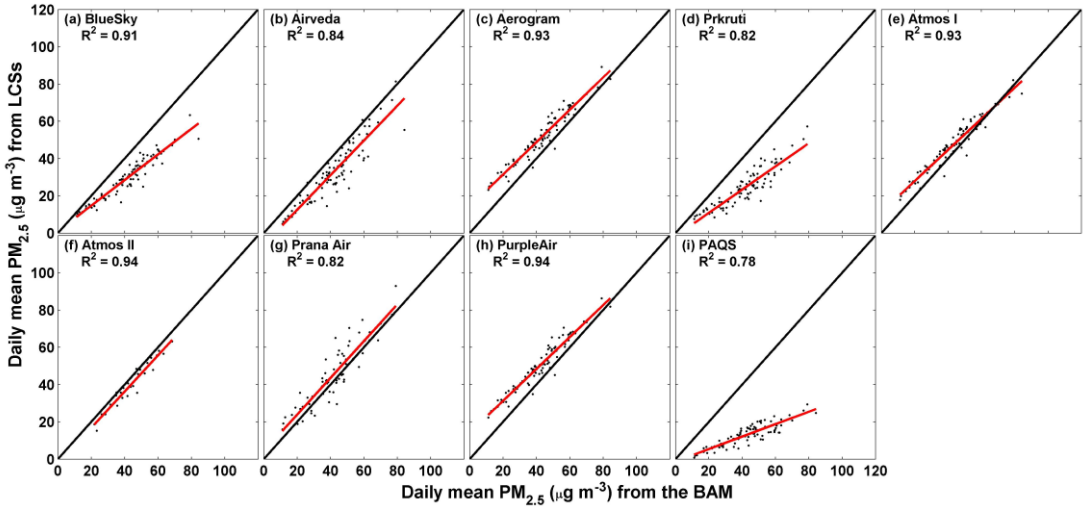


Figure A2: Scatter plots of daily mean LCSs and BAM $PM_{2.5}$ for winter (DJF). Red and black lines indicate linear fit and 1:1 lines, respectively.

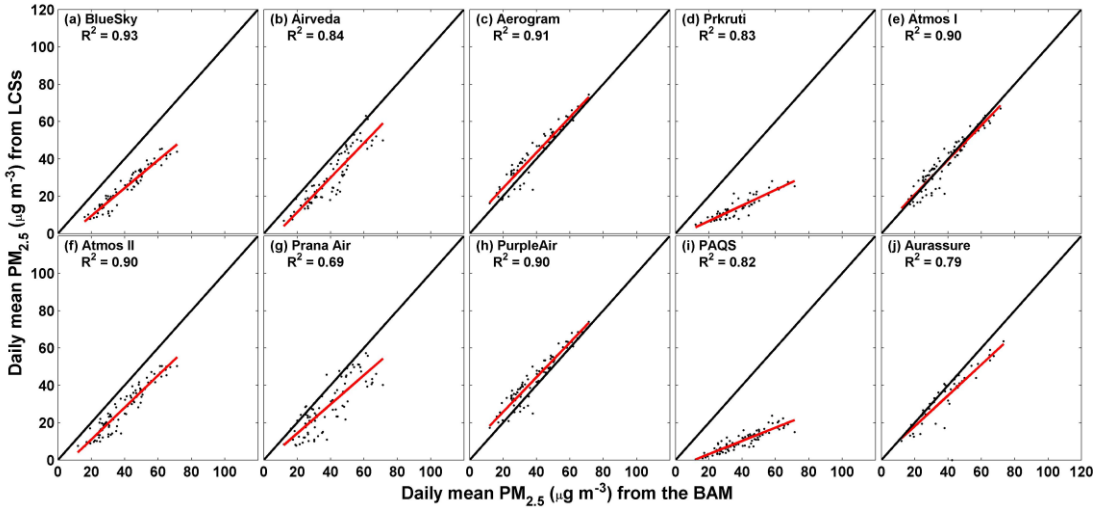


Figure A3: Scatter plots of daily mean LCSs and BAM $PM_{2.5}$ for summer (MAM). Red and black lines indicate linear fit and 1:1 lines, respectively.

Table A4: Season-wise performance statistics of daily mean LCSs $PM_{2.5}$
(N represents the number of paired data points)

Sensor	N		R^2		RMSE (µg m^{-3})		NRMSE	
	DJF	MAM	DJF	MAM	DJF	MAM	DJF	MAM
BlueSky	89	76	0.91	0.93	13.6	16.5	0.32	0.40
Airveda	89	80	0.84	0.84	11.1	11.5	0.26	0.28
Aerogram	89	75	0.93	0.91	9.4	5.5	0.22	0.14
Prkruti	88	68	0.82	0.83	18.8	24.3	0.45	0.67
Atmos I	89	91	0.93	0.90	6.1	4.4	0.14	0.11
Atmos II	29	91	0.94	0.90	4.9	$1_{2.5}$	0.11	0.32
Prana Air	63	82	0.82	0.69	7.9	12.8	0.19	0.33
PurpleAir (cf_atm)	89	87	0.94	0.90	9.0	6.1	0.21	0.16
PAQS	89	91	0.78	0.82	31.4	30.4	0.74	0.78
Aurassure	NA	53	NA	0.79	NA	8.7	NA	0.23

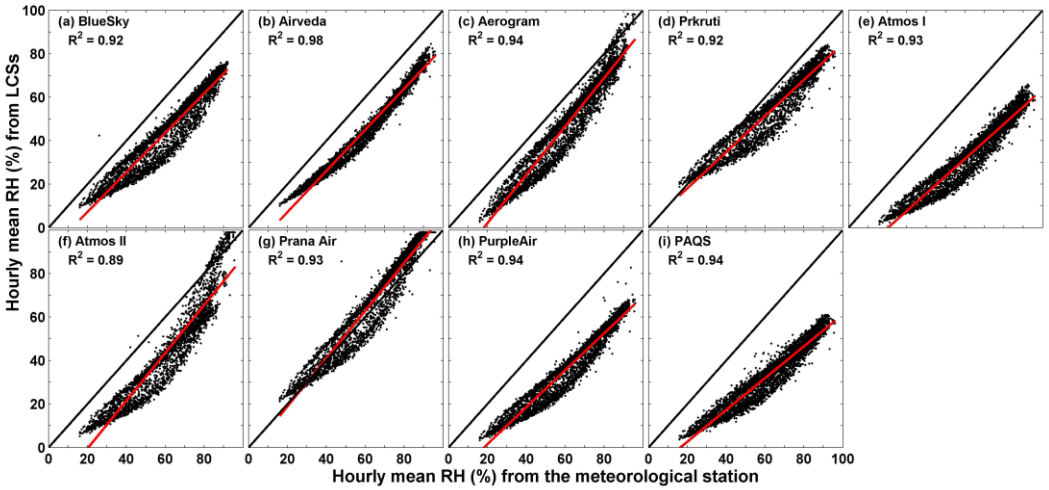


Figure A4: Scatter plots of hourly LCSs and the meteorological station RH (%). Red and black lines indicate linear fit and 1:1 lines, respectively.

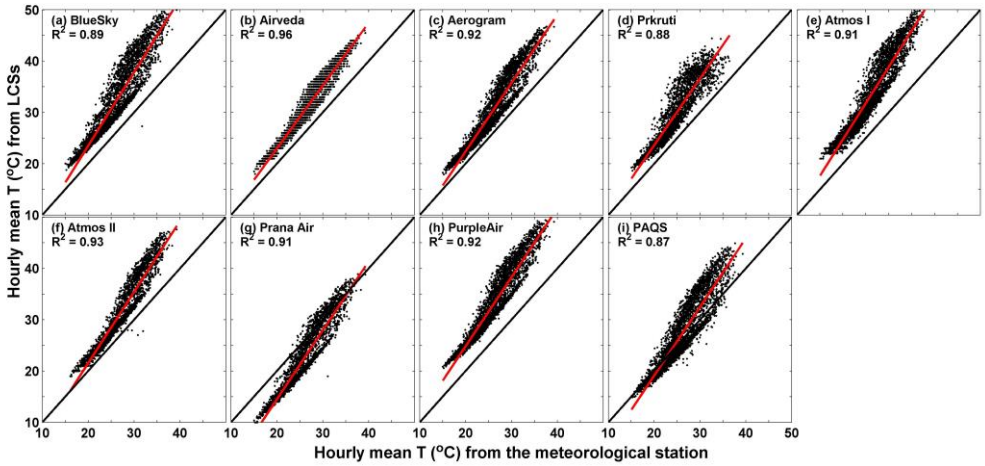


Figure A5: Scatter plots of hourly LCSs and the meteorological station temperature (T, °C). Red and black lines indicate linear fit and 1:1 lines, respectively.

Table A5: Performance statistics of hourly LCSs RH

Sensor	N	R^2	RMSE (%)	NRMSE
BlueSky	3890	0.92	17.3	0.28
Airveda	4038	0.98	15.4	0.24
Aerogram	3906	0.94	14.3	0.23
Prkruti	3148	0.92	11.0	0.17
Atmos I	4319	0.93	29.6	0.46
Atmos II	2844	0.89	18.1	0.30
Prana Air	3157	0.93	6.7	0.11
PurpleAir (cf_atm)	4204	0.94	25.4	0.40
PAQS	4295	0.94	29.7	0.47

Table A6: Performance statistics of hourly LCSs temperature (T)

Sensor	N	R^2	RMSE (^oC)	NRMSE
BlueSky	3890	0.89	6.5	0.26
Airveda	4038	0.96	4.4	0.18
Aerogram	3906	0.92	4.8	0.19
Prkruti	3148	0.88	5.4	0.23
Atmos I	4319	0.91	7.4	0.30
Atmos II	2844	0.93	4.7	0.18
Prana Air	3157	0.91	4.5	0.18
PurpleAir (cf_atm)	4204	0.92	7.0	0.28
PAQS	4295	0.87	3.0	0.12

Table A7: Regression coefficients of PA $PM_{2.5}$ monthly correction models

Months	Dec 2021	Jan 2022	Feb 2022	Mar 2022	Apr 2022	May 2022
Model R^2	0.95	0.85	0.94	0.93	0.91	0.57
Intercept (p-value)	-19.35 (0)	-6.16 (1e-4)	-14.93 (0)	-10.59 (0)	-14.84 (0)	-8.85 (3e-5)
BAM coefficient (p-value)	1.54 (0)	1.23 (0)	1.45 (0)	1.47 (0)	1.60 (0)	1.13 (0)
RH coefficient (p-value)	0.57 (0)	0.53 (0)	0.71 (0)	0.52 (0)	0.40 (0)	0.25 (0)

Table A8: Performance metrics of month-specific multivariate calibration models. N represents the number of paired data points.

Months	N		R^2		RMSE (µg m^{-3})		NRMSE	
	Hourly	Daily	Hourly	Daily	Hourly	Daily	Hourly	Daily
Dec 2021	729	31	0.95	0.98	5.9	3.0	0.14	0.07
Jan 2022	706	30	0.84	0.90	10.4	5.4	0.25	0.13
Feb 2022	651	28	0.90	0.96	5.6	2.9	0.13	0.07
Mar 2022	649	28	0.92	0.98	5.6	2.3	0.12	0.05
Apr 2022	697	30	0.91	0.97	4.7	2.1	0.12	0.06
May 2022	654	28	0.58	0.51	9.4	7.7	0.32	0.26

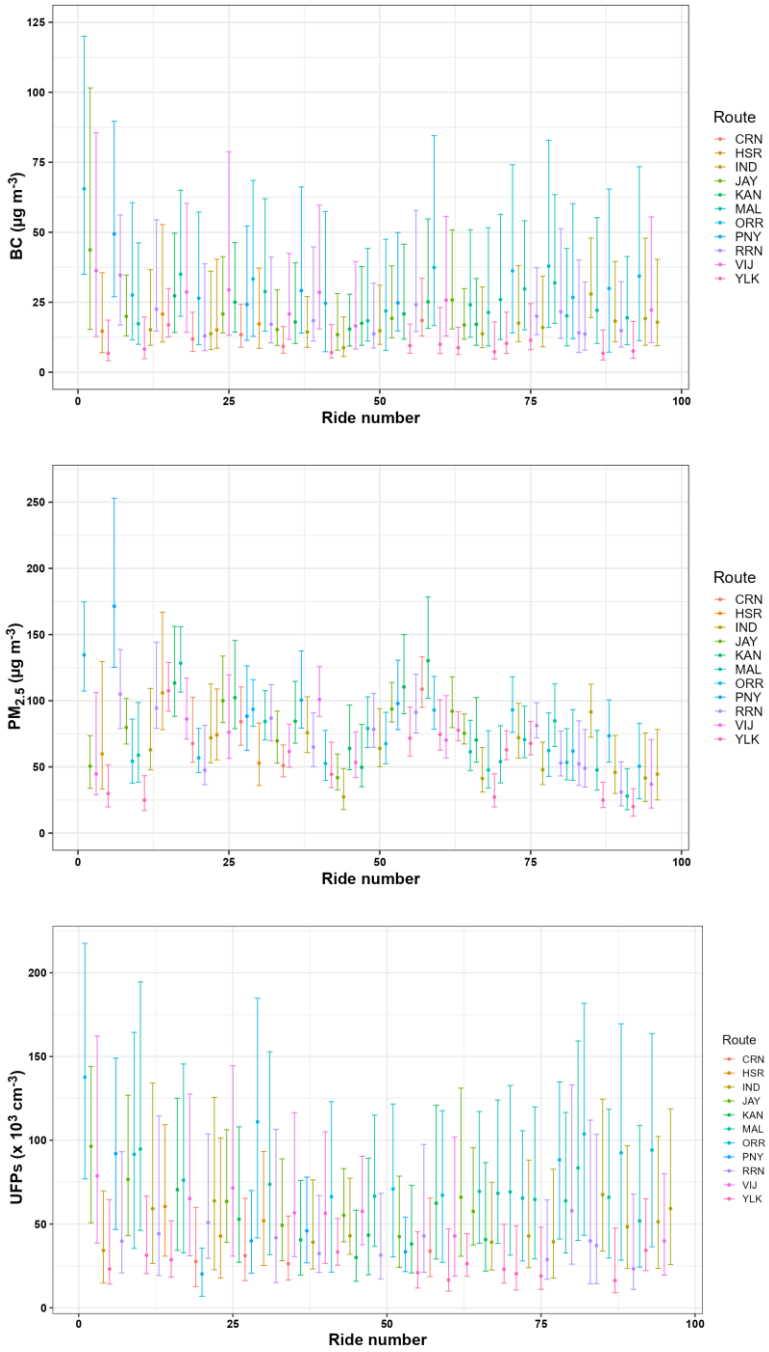


Figure A6: Ride-to-ride variations in on-road gridded pollutant levels. Colours indicate different routes. The dot indicates the median, and whiskers indicate the inter-quartile range.

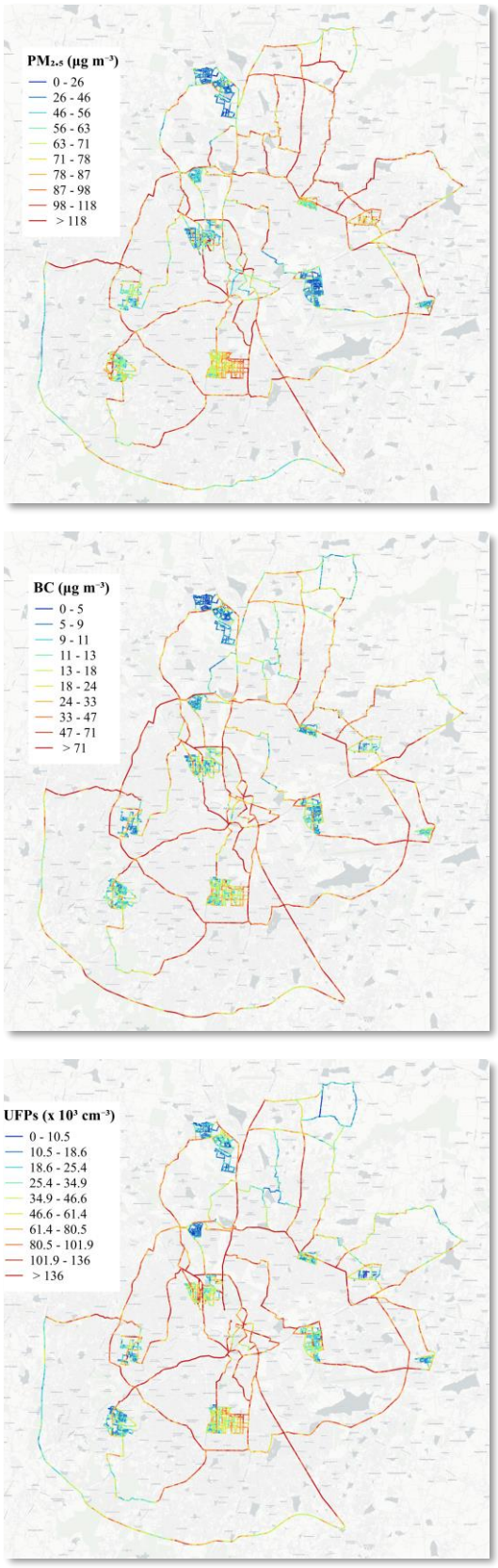


Figure A7: Route maps of on-road $PM_{2.5}$, BC, and UFPs derived using data from eight repeat rides

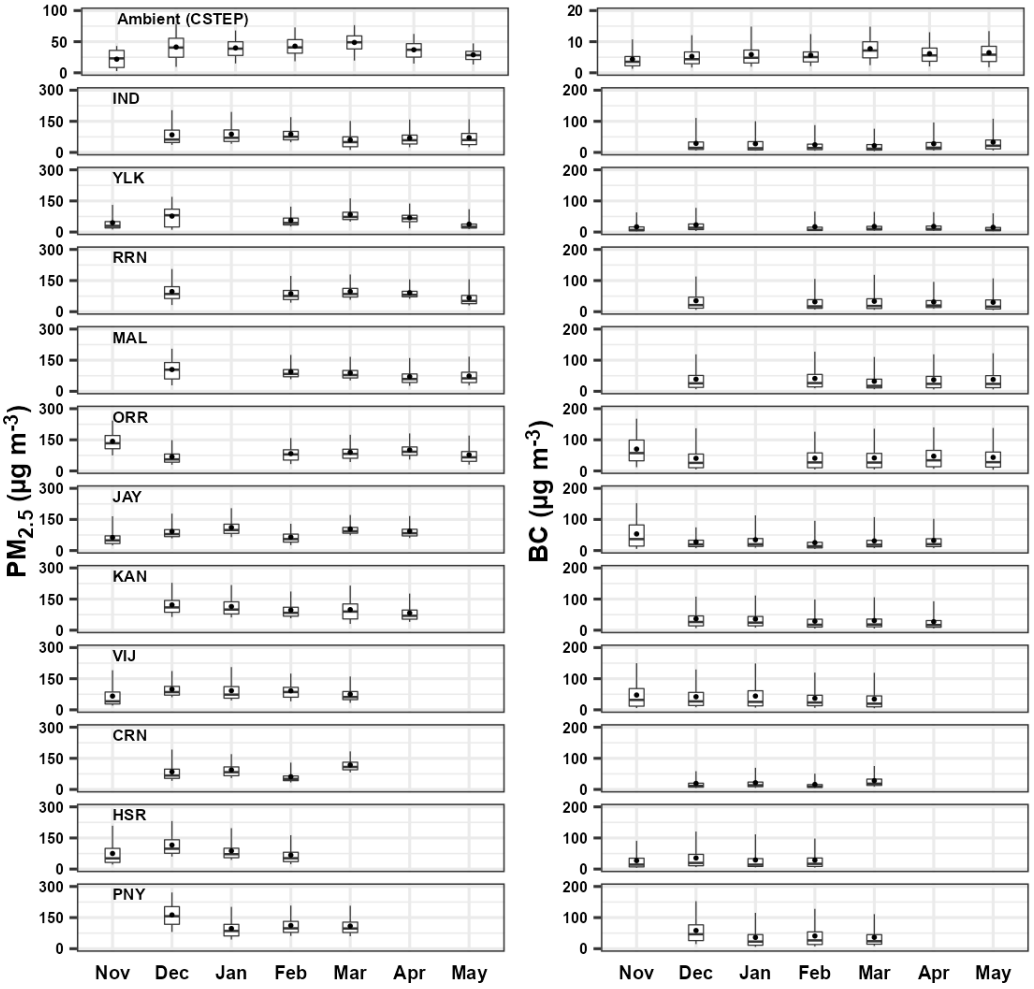


Figure A8: Route-wise comparison of ambient and on-road $PM_{2.5}$ and BC

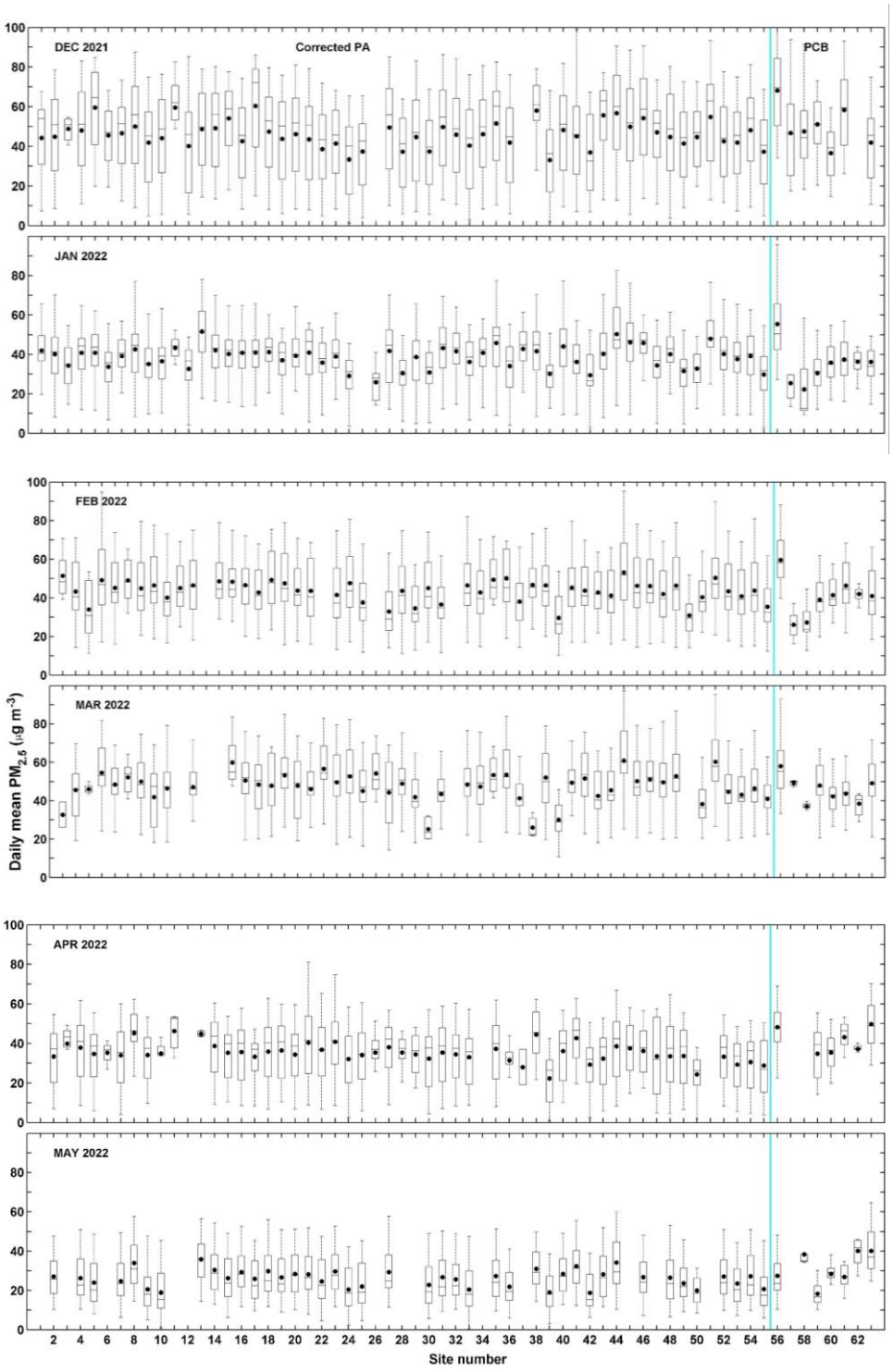


Figure A9: Month- and site-wise daily mean $PM_{2.5}$ distribution

10. Appendix B

As part of the project, we developed a desktop application (in R Shiny) to visualise and manage sensor history. Currently, the application has all the basic functionalities. It is useful in sharing details of installation and history of sensors among team members and external people, if required. It tracks the status of installed sensors based on an input Google spreadsheet. The spreadsheet should contain relevant details of sensors. The application gives the flexibility to view information about a particular sensor or installation location filtered by location or sensor ID. The following screenshots navigate through various tabs and options available in the application.

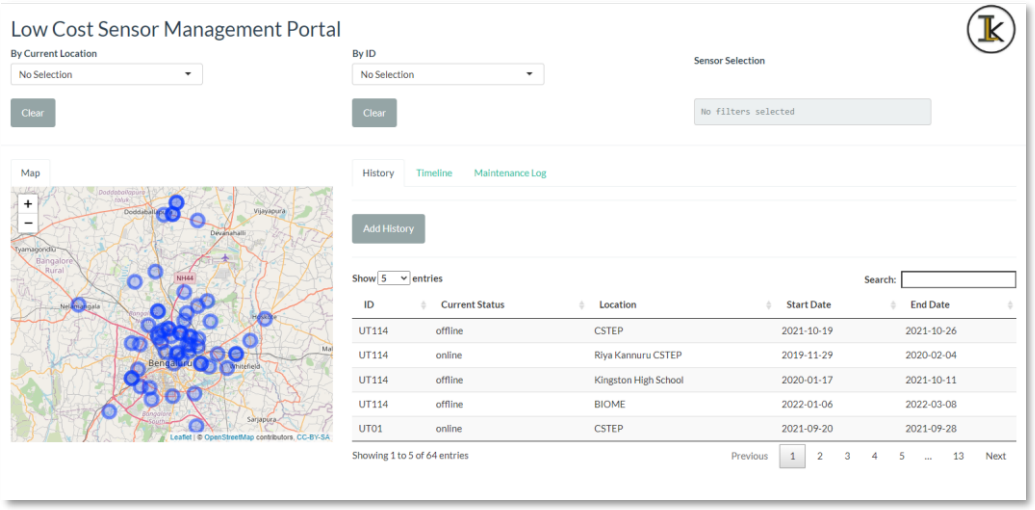


Figure B1: Home screen of the application

On the home screen, (i) a list of tabs, (ii) a Google map with geographical locations of sensors, (iii) a table with sensor details, (iv) a search bar, and (v) two filter drop-down menus (by sensor id and name) are available.

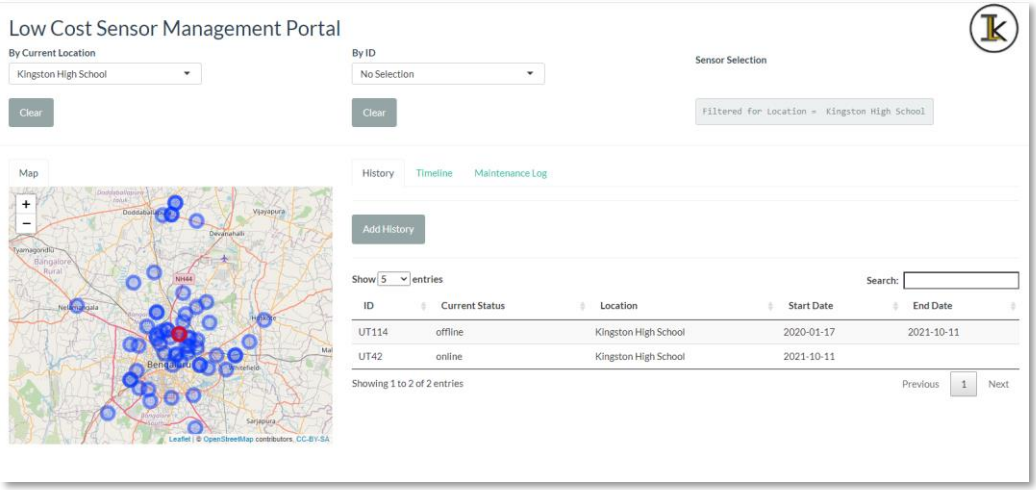


Figure B2: Home screen after filtering by a location name

If a location name is selected, its geographical location will be highlighted (in red on the map) and details of sensors installed in that location will be listed with their start and end dates.

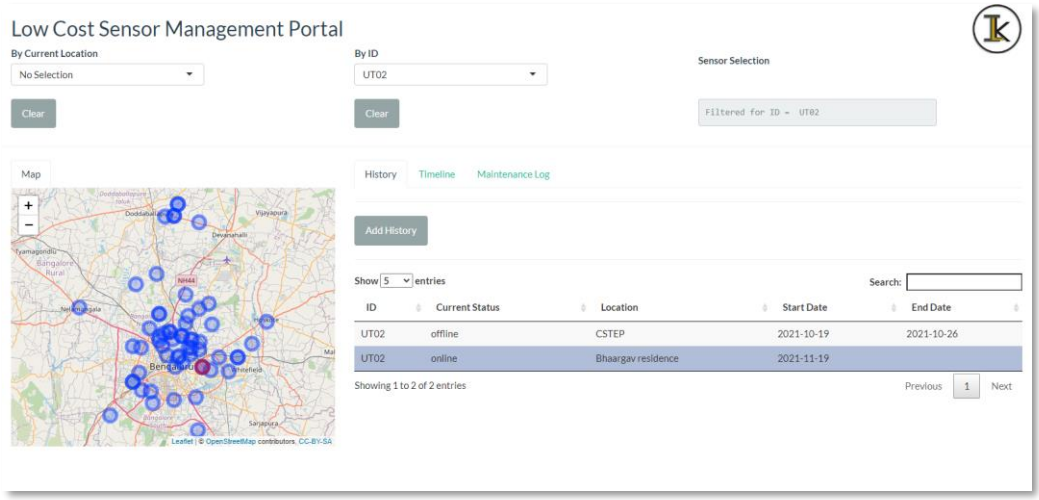


Figure B3: Home screen after filtering by a sensor ID

If a sensor is filtered by its ID, current and historical locations of its installation and start and end dates will be listed.

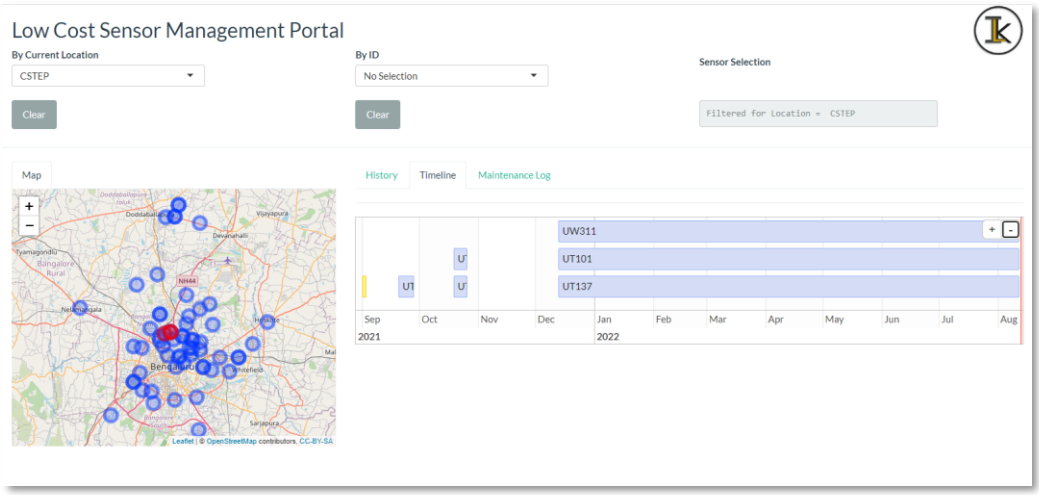


Figure B4: Timeline tab

If a location name is selected, the Timeline tab provides information on the active period of various sensors in that location in a single plot. Zoom in/out option for the calendar is also available.

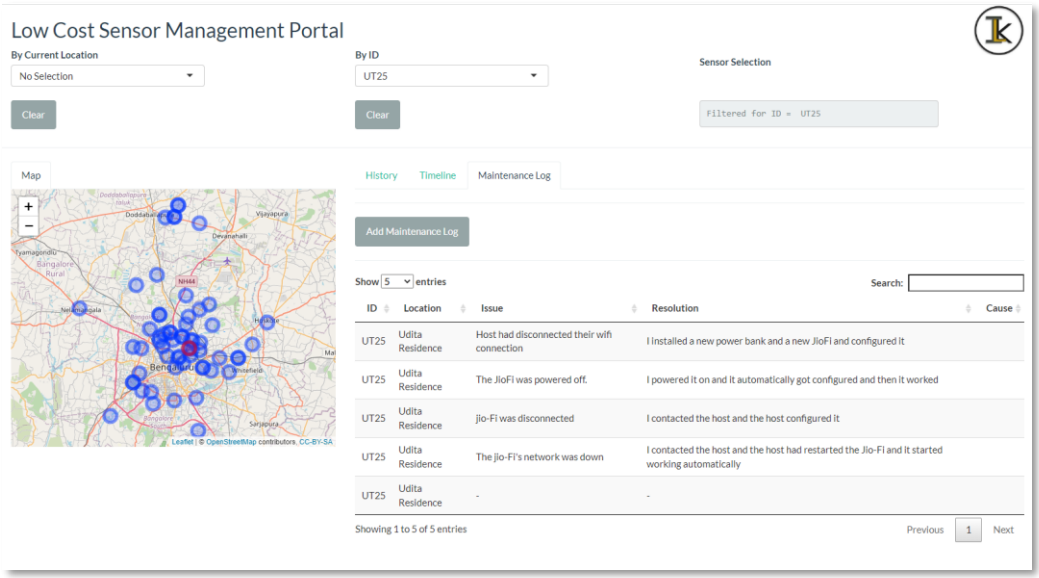


Figure B5: Maintenance Log tab

The Maintenance Log tab provides information on troubleshooting activities carried out on a sensor in that location.

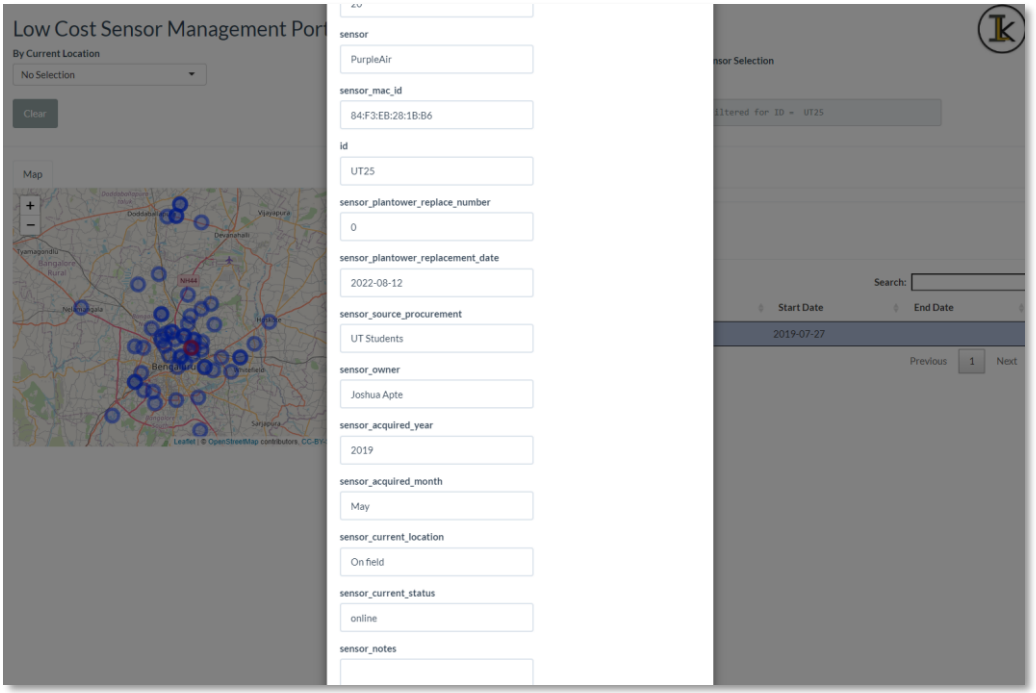
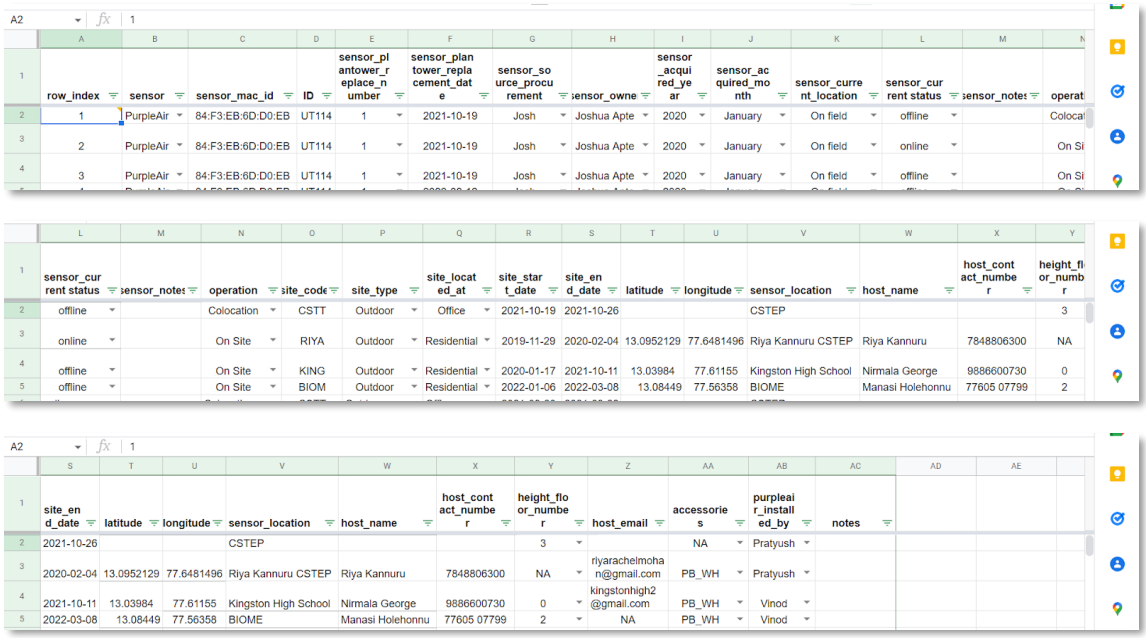


Figure B6: Sensor details

Sensor details will be displayed when the sensor ID field is selected.

This application has scope for further development and currently operates on the input spreadsheet. It can be linked to the sensor website/dashboard to check current status, download data, plot, visualise, compare, and derive meaningful statistics. Data security and login features can also be added.



row_index	sensor	sensor_mac_id	ID	sensor_plantower_replacement_number	sensor_plantower_replacement_date	sensor_source_procurement	sensor_owner	sensor_acquired_year	sensor_acquired_month	sensor_current_location	sensor_current status	sensor_notes	operat
1	PurpleAir	84:F3:EB:6D:D0:EB	UT114	1	2021-10-19	Josh	Joshua Apte	2020	January	On field	offline		Colocat
2	PurpleAir	84:F3:EB:6D:D0:EB	UT114	1	2021-10-19	Josh	Joshua Apte	2020	January	On field	online		On Si
3	PurpleAir	84:F3:EB:6D:D0:EB	UT114	1	2021-10-19	Josh	Joshua Apte	2020	January	On field	offline		On Si

sensor_current status	sensor_notes	operation	site_code	site_type	site_located_at	site_start_date	site_end_date	latitude	longitude	sensor_location	host_name	host_contact_number	height_floor_number
offline		Colocation	CSTT	Outdoor	Office	2021-10-19	2021-10-26			CSTEP			3
online		On Site	RIYA	Outdoor	Residential	2019-11-29	2020-02-04	13.0952129	77.6481496	Riya Kannuru CSTEP	Riya Kannuru	7848806300	NA
offline		On Site	KING	Outdoor	Residential	2020-01-17	2021-10-11	13.03984	77.61155	Kingston High School	Nirmala George	9886600730	0
offline		On Site	BIOM	Outdoor	Residential	2022-01-06	2022-03-08	13.08449	77.56358	BIOME	Manasi Holehonnu	77605 07799	2

site_end_date	latitude	longitude	sensor_location	host_name	host_contact_number	height_floor_number	host_email	accessories	purpleair_installed_by	notes
2021-10-26			CSTEP			3		NA	Pratyush	
2020-02-04	13.0952129	77.6481496	Riya Kannuru CSTEP	Riya Kannuru	7848806300	NA	riyarachelmohan@gmail.com	PB_WH	Pratyush	
2021-10-11	13.03984	77.61155	Kingston High School	Nirmala George	9886600730	0	kingstonhigh2@gmail.com	PB_WH	Vinod	
2022-03-08	13.08449	77.56358	BIOME	Manasi Holehonnu	77605 07799	2	NA	PB_WH	Vinod	

Figure B7: Snapshots of input spreadsheets

CENTER FOR STUDY OF SCIENCE, TECHNOLOGY & POLICY

Bengaluru
#18 & 19, 10th Cross, Mayura Street,
Papanna Layout, Nagashettyhalli (RMV II Stage),
Bengaluru-560094, Karnataka, India

Noida
1st Floor, Tower-A, Smartworks Corporate Park, Sector-125,
Noida-201303, Uttar Pradesh, India

www.cstep.in | +91-8066902500 | cpe@cstep.in | @cstep_India