



Satellite-Based Mapping and the Quantification of PM_{2.5} in India



Satellite-Based Mapping and the Quantification of PM_{2.5} in India

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

February 2022

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

Designed and edited by CSTEP

Disclaimer

While every effort has been made for the correctness of data/information used in this report, neither CSTEP nor the authors accept any legal liability for the accuracy or inferences for the material contained in this working series and for any consequences arising from the use of this material.

© 2022 Center for Study of Science, Technology and Policy (CSTEP)

Any reproduction in full or part of this publication must mention the title and/or citation, which is provided below. Due credit must be provided regarding the copyright owners of this product.

Contributors: Dr Sreekanth Vakacherla and Dr Padmavati Kulkarni

(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: CSTEP. (2022). Satellite-based mapping and the quantification of PM_{2.5} in India. (CSTEP-RR-2022-07).

February 2022

Center for Study of Science, Technology and Policy

Bengaluru

18, 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)

Tel.: +91 (80) 6690 2500

Email: cpe@cstep.in

Acknowledgements

We would like to take this opportunity to thank our funders, collaborators, and colleagues for their continued support and guidance throughout the project.

First of all, we thank Environmental Defense Fund (EDF) for providing financial support to CSTEP for conducting air pollution studies in Bengaluru.

We convey our sincere gratitude to Dr Hrishikesh Chandra Gautam for his role in building artificial intelligence-based models that significantly helped further this study.

We thank our internal reviewers Mr Ashish Srivastava and Mr Pareexit Chauhan and our external reviewers Dr Santu Ghosh, St. John's Medical College, Bengaluru, and Dr Naveen Puttaswamy, Sri Ramachandra Institute of Higher Education and Research, Chennai, for their critical inputs on the project.

We also take this opportunity to acknowledge the CSTEP leadership—Dr Jai Asundi, Mr Priyavrat Bhati, and Dr Pratima Singh—for their valuable support and guidance throughout.

We thank the Communications and Policy Engagement (CPE) team of CSTEP for their editorial guidance.

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

MODIS satellite data have been obtained from the National Aeronautics and Space Administration (NASA) Landsweb data archive. We would like to thank the MODIS team for developing the aerosol products.

ERA-5 data have been obtained from the Climate Data Store (CDS) of the Copernicus Climate Change Service (C3S). We acknowledge the European Centre for Medium-Range Weather Forecasts (ECMWF), CDS, and C3S teams for the data.

Global Human Settlement Layer Settlement Model (GHSL-SMOD) data have been obtained by the European Commission. We would like to thank the GHSL team for developing the population settlement data.

We also acknowledge the National Geophysical Data Center (NGDC) and the National Oceanic and Atmospheric Administration (NOAA) for the ETOPO2 elevation data used in the present study.

We are grateful to the Indo-Gangetic Plains – Center for Air Research and Education (IGP-CARE) team for providing the rural air pollution data used in the study.

Executive Summary

Regulatory air pollution monitoring in India is mostly limited to urban areas. Without a dense network of monitors, it is difficult to capture the fine spatial variations of PM_{2.5}, one of the major pollutants with severe implications for human health. Using satellite-based products to estimate PM_{2.5} can help generate high-resolution gridded spatial maps at a significantly lower cost. However, there are challenges and limitations in converting the satellite columnar aerosol optical depth (AOD) into surface PM_{2.5}. Major limitations include the non-availability of satellite AOD during cloudy days and the lack of non-urban PM_{2.5} measurements.

In this study, the daily mean PM_{2.5} was estimated and spatial maps (of 1 km spatial resolution) were generated using Moderate Resolution Imaging Spectroradiometer (MODIS) AOD for the calendar year 2019 across select Indian regions. The study regions included the urban, peri-urban, and rural regions of Delhi-National Capital Region (NCR), Kanpur, and Bengaluru. An advanced statistical model was trained using open-access data sets (satellite, regulatory ground-based PM_{2.5}, reanalysis meteorology) to estimate the daily mean PM_{2.5}. Annual and seasonal maps of PM_{2.5} were generated and a hotspot analysis was performed to identify spatial clusters of high PM_{2.5} grids within the study regions. Spatial gradients in PM_{2.5} were studied to understand the rural, peri-urban, and urban contrast in pollution levels.

The annual mean PM_{2.5} values for Delhi-NCR, Kanpur, and Bengaluru ranged between 80 and 130 $\mu\text{g m}^{-3}$, 95 and 130 $\mu\text{g m}^{-3}$, and 35 and 55 $\mu\text{g m}^{-3}$, respectively. The annual mean PM_{2.5} in all the spatial grids of Delhi-NCR and Kanpur exceeded the national annual standard (40 $\mu\text{g m}^{-3}$). Seasonally, monsoon months recorded the lowest PM_{2.5} and winter months the highest PM_{2.5} across all the study regions. PM_{2.5} hotspots were identified on annual and seasonal scales. Not surprisingly, most of the hotspots were clustered within the urban regions of Delhi, Kanpur, and Bengaluru.

Most of the Delhi-National Capital Territory and its surrounding areas were identified as hotspot regions. Similarly, most of the urban zones in Kanpur were identified as hotspots. In Bengaluru, hotspots were observed in parts of Bommanahalli, Dasarahalli, Rajarajeshwari Nagar, South and West zones, and the Doddaballapura Taluk (rural). Weak spatial gradients in the annual and seasonal mean PM_{2.5} were observed across all of the study regions. The annual mean PM_{2.5} in rural Delhi-NCR was observed as $101 \pm 6.5 \mu\text{g m}^{-3}$ (mean \pm standard deviation), which is comparable to the annual mean ($109 \pm 9.4 \mu\text{g m}^{-3}$) in urban Delhi-NCR. Even the uninhabited areas of Delhi-NCR recorded an annual mean of PM_{2.5} $100 \pm 6.6 \mu\text{g m}^{-3}$. Similarly, for Kanpur, the urban and rural areas recorded annual means of PM_{2.5} $114 \pm 6.4 \mu\text{g m}^{-3}$ and $108 \pm 4.6 \mu\text{g m}^{-3}$, respectively. Bengaluru too exhibited little ($\sim 2 \mu\text{g m}^{-3}$) rural-urban contrast.

The study provides extensive PM_{2.5} statistics, which could be highly useful for policymakers, researchers, and citizen scientists. Major policy recommendations include establishing representative rural and peri-urban regulatory pollution monitoring stations and identifying pollution sources in non-urban areas and unorganised sectors.

Contents

1. Introduction.....	15
1.1 Background.....	15
1.2 Literature review	15
1.3 Need for the study.....	16
1.4 Challenges	17
2. Objectives of the Study.....	18
3. Study Areas and Study Period	19
4. Data Sets.....	21
4.1 Satellite Aerosol Optical Depth (AOD).....	21
4.1.1 MODIS.....	21
4.1.2 MAIAC.....	21
4.2 PM _{2.5}	22
4.3 Meteorological data.....	22
4.4. Auxiliary data.....	23
5. Data Handling	25
5.1 AOD.....	25
5.2 PM _{2.5}	25
5.3 Meteorological parameters	25
5.3.1 Derived variables	25
5.3.2 Spatial interpolation	26
5.4 NDVI	26
6. Methodology	27
6.1 Model selection.....	27
6.2 Random forest (RF) model.....	28
6.3 Model validation.....	28
6.4 Monitoring stations.....	28
6.5 Hotspot analysis	29
6.6 Rural, peri-urban, urban, and uninhabited pixel classification	29
7. Results	31
7.1 Spatio-temporal variations in the daily mean PM _{2.5} and AOD.....	31
7.2 Model evaluation.....	33
7.3 Spatial maps of PM _{2.5}	35
7.3.1 Delhi-NCR region.....	35
7.3.2 Kanpur region.....	36

7.3.3 Bengaluru region.....	37
7.3.4 Urban areas of Delhi, Kanpur, and Bengaluru.....	39
7.4 PM _{2.5} hotspots.....	47
7.4.1 Delhi-NCR region.....	47
7.4.2 Kanpur region.....	48
7.4.3 Bengaluru region.....	50
7.5 Spatial gradients.....	51
7.5.1 Delhi-NCR region.....	51
7.5.2 Kanpur region.....	52
7.5.3 Bengaluru region.....	53
8. Conclusions	57
8.1 Limitations.....	58
9. Way Forward	59
9.1 Future research recommendations	59
9.2 Policy recommendations	59
10. References.....	60
11. Appendix A: Supplemental Figures.....	65
12. Appendix B: Satellite AOD-Derived PM _{2.5} Maps of Indore City	77

Figures

Figure 1: Delhi-NCR (first panel), Kanpur (second panel), and Bengaluru (third panel)	19
Figure 2: Geographical locations of the 89 CAAQMS overlaid on the state boundaries of (a) Karnataka and (b) UP. The blue line in panel (b) represents Delhi-NCR.....	29
Figure 3: Spatio-temporal variations in the daily mean PM _{2.5} . Individual stations' data from each region were combined.	31
Figure 4: Spatio-temporal variations in the daily mean AOD.....	32
Figure 5: Density scatter plots depicting the model-predicted daily mean PM _{2.5} and measured daily mean PM _{2.5} (the Delhi-NCR and Kanpur model).....	33
Figure 6: Density scatter plots depicting the model-predicted daily mean PM _{2.5} and measured daily mean PM _{2.5} (the Bengaluru model)	34
Figure 7: The spatial map of the annual mean PM _{2.5} over Delhi-NCR and its frequency distribution.....	35
Figure 8: The seasonal mean PM _{2.5} maps over Delhi-NCR	36
Figure 9: The spatial map of the annual mean PM _{2.5} over the Kanpur region and its frequency distribution	37
Figure 10: The seasonal mean PM _{2.5} maps over the Kanpur region.....	37
Figure 11: The spatial map of the annual mean PM _{2.5} over the Bengaluru region and its frequency distribution	38
Figure 12: The seasonal mean PM _{2.5} maps over the Bengaluru region.....	39
Figure 13: The spatial map of the annual mean PM _{2.5} over Delhi-NCT region with overlaid revenue district and ward boundaries.....	40
Figure 14: Spatial maps of the seasonal mean PM _{2.5} over Delhi-NCT	40
Figure 15: The spatial map of the annual mean PM _{2.5} over Kanpur urban (zone and ward boundaries overlaid)	43
Figure 16: Spatial maps of the seasonal mean PM _{2.5} over Kanpur urban	43
Figure 17: The spatial map of the annual mean PM _{2.5} over Bengaluru urban (zone and ward boundaries overlaid)	45
Figure 18: Spatial maps of the seasonal mean PM _{2.5} over Bengaluru urban	45
Figure 19: Annual hotspot analysis for Delhi-NCR. The right panel corresponds to the GHSL-SMOD settlement classification.....	47
Figure 20: Seasonal hotspot analysis for Delhi-NCR	48
Figure 21: Annual hotspot analysis for the Kanpur region	49
Figure 22: Seasonal hotspot analysis for the Kanpur region.....	49
Figure 23: Annual hotspot analysis for the Bengaluru region	50
Figure 24: Seasonal hotspot analysis for the Bengaluru region.....	51
Figure 25: Variations in the annual mean PM _{2.5} for Delhi-NCR for the identified settlement classes: uninhabited, rural, peri-urban, and urban	52

Figure 26: Variations in the seasonal mean PM _{2.5} for Delhi-NCR for the identified settlement classes.....	52
Figure 27: Variations in the annual mean PM _{2.5} for the Kanpur region for the identified settlement classes	53
Figure 28: Variations in the seasonal mean PM _{2.5} of the Kanpur region for the identified settlement classes	53
Figure 29: Variations in the annual mean PM _{2.5} of the Bengaluru region for the identified settlement classes.....	54
Figure 30: Variations in the seasonal mean PM _{2.5} of the Bengaluru region for the identified settlement classes.....	54
Figure A1: AOD spatial distribution before and after imputation for a typical winter day	65
Figure A2: Typical raw and quality-checked daily mean PM _{2.5} time series	65
Figure A3: Bilinear interpolation	66
Figure A4: The cubic-spline temporal interpolation of NDVI.....	66
Figure A5: Spatio-temporal variations in the daily mean NDVI	67
Figure A6: Spatio-temporal variations in the daily mean PBLH	67
Figure A7: Spatio-temporal variations in the daily mean surface pressure	67
Figure A8: Spatio-temporal variations in the daily mean relative humidity.....	68
Figure A9: Spatio-temporal variations in the daily mean surface temperature	68
Figure A10: Spatio-temporal variations in the wind speed	68
Figure A11: Density scatter plots depicting the model-predicted monthly mean PM _{2.5} and measured monthly mean PM _{2.5} . Monthly mean values were derived by aggregating daily mean values to the calendar month.	69
Figure A12: Density scatter plots depicting the model-predicted daily mean PM _{2.5} and measured daily mean PM _{2.5} (the Delhi-NCR and Kanpur model). Here, the LME model was configured with the site- specific random intercept, in addition to the day-specific random slope and intercept.....	69
Figure A13: Density scatter plots depicting the model-predicted daily mean PM _{2.5} and measured daily mean PM _{2.5} (the Bengaluru model). Here, the LME model was configured with the site-specific random intercept, in addition to the day-specific random slope and intercept.....	70
Figure A14: The performance of the random forest model (10-fold CV) for Delhi-NCR and Kanpur	70
Figure A15: The comparison of the LME-predicted and measured PM _{2.5} for a rural location in UP	71
Figure B1: Density scatter plots depicting the model-predicted daily mean PM _{2.5} and measured daily mean PM _{2.5} (the Indore model).....	78
Figure B2: The spatial map of the annual mean PM _{2.5} over Indore for the year 2019 and surrounding regions and their frequency distribution	78
Figure B3: The seasonal mean maps of PM _{2.5} over Indore. The solid black line represents the Indore municipal area	79

Tables

Table 1: The list of shortlisted predictors used for model building.	23
Table 2: Season-wise daily mean PM _{2.5} statistics. IQR and SD represent the interquartile range and standard deviation, respectively.	32
Table 3: Season-wise AOD statistics. IQR and SD represent the interquartile range and standard deviation, respectively.	33
Table 4: Annual and seasonal mean PM _{2.5} statistics. IQR and SD represent the interquartile range and standard deviation, respectively.	39
Table 5: Annual and seasonal statistics of PM _{2.5} for various Delhi-NCT revenue districts. IQR and SD represent the interquartile range and standard deviation, respectively.	40
Table 6: Annual and seasonal statistics of PM _{2.5} for various Kanpur urban zones. IQR and SD represent the interquartile range and standard deviation, respectively.	43
Table 7: Annual and seasonal statistics of PM _{2.5} for various Bengaluru urban zones. IQR and SD represent the interquartile range and standard deviation, respectively.	45
Table 8: The annual and season-wise mean PM _{2.5} statistics for different settlement classes of study regions. IQR and SD represent the interquartile range and standard deviation, respectively.	55
Table A1: Details of the CAAAQMS stations from which the PM _{2.5} data are used in this study along with PM _{2.5} and MODIS-AOD annual statistics. N indicates the number of daily mean data points available in the study year. SD represents the standard deviation.	72
Table B1: CAAQMSs from which PM _{2.5} data is used for building the Indore model. SD and N represent the standard deviation and the number of daily mean data points.	77
Table B2: The seasonal mean PM _{2.5} statistics for Indore. SD represents the standard deviation.	79

Abbreviations and Acronyms

µg	Microgram
µm	Micrometre
AERONET	Aerosol Robotic Network
AOD	Aerosol optical depth
BAM	Beta attenuation monitor
BC	Black carbon
BRDF	Bidirectional reflectance distribution function
CAAQMS	Continuous Ambient Air Quality Monitoring Stations
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder
cm	Centimetre
CPCB	Central Pollution Control Board
CSTEP	Center for Study of Science, Technology and Policy
CV	Cross validation
CWV	Columnar water vapour
DB	Deep Blue
DEM	Digital Elevation Model
DT	Dark Target
DPT	Dew point temperature
ECMWF	European Centre for Medium-Range Weather Forecasts
ELV	Elevation
ERA	European Centre for Medium-Range Weather Forecasts Reanalysis
FEMs	Federal equivalent methods
FRM	Federal reference method
GAM	Generalized Additive Model
GBD	Global Burden of Disease
GHSL-SMOD	Global Human Settlement Layer Settlement Model
GWR	Geographically weighted regression
IGP	Indo-Gangetic Plains
IQR	Inter-quartile range
km	Kilometre
kmph	Kilometres per hour
LME	Linear mixed-effects
LMICs	Low- and middle-income countries
LOOCV	Leave-one-out cross-validation
LPG	Liquefied petroleum gas
LT	Local time

LUT	Lookup table
MAB	Mean absolute bias
MAIAC	Multi-Angle Implementation of Atmospheric Correction
MERRA-2	Modern-Era Retrospective Analysis for Research and Applications Version 2
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MoEFCC	Ministry of Environment, Forest and Climate Change
NAAQS	National Ambient Air Quality Standards
NASA	National Aeronautics and Space Administration
NCAP	National Clean Air Programme
NCR	National Capital Region
NCT	National Capital Territory
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
NOAA	National Oceanic and Atmospheric Administration
NRMSE	Normalised root mean square error
OMI	Ozone Monitoring Instrument
PBLH	Planetary boundary layer height
PM ₁₀	Particulate matter with a diameter of 10 micrometres or less
PM _{2.5}	Particulate matter with a diameter of 2.5 micrometres or less
QA	Quality assurance
RF	Random forest
RH	Relative humidity
RMSE	Root mean square error
SD	Standard deviation
SPCB	State Pollution Control Board
ST	Surface air temperature
TEOM	Tapered element oscillating microbalance
UP	Uttar Pradesh
USEPA	United States Environmental Protection Agency
VIIRS	Visible Infrared Imaging Radiometer Suite

013 PM2.5



1. Introduction

1.1 Background

Globally, air pollution was found to be the fourth leading risk factor for premature mortality in 2019 (Murray et al., 2020). $PM_{2.5}$, particulate matter (PM) with a size smaller than 2.5 microns (μm), is one of the major health-damaging air pollutants that low- and middle-income countries (LMICs) are battling with at the moment. $PM_{2.5}$ is known to impact climate, environment, public health, and the economy. Detrimental health impacts (such as cardiovascular, respiratory, and neurodegenerative diseases) of $PM_{2.5}$ are well understood and documented (Kim et al., 2015; Schraufnagel et al., 2019). According to the Health Effects Institute (2020), developing countries such as Nigeria, Bangladesh, India, and Pakistan have the highest $PM_{2.5}$ exposures, and these exposures are increasing. In India, around one million deaths in 2019 were attributed to ambient $PM_{2.5}$.

Conventionally, ambient $PM_{2.5}$ is quantified using instruments employing a variety of measurement principles. The gravimetric method is considered the gold standard for the measurement of $PM_{2.5}$ and was designated as the Federal Reference Method (FRM) by the United States Environmental Protection Agency (USEPA). This method provides time-integrated average estimates of $PM_{2.5}$ based on the weighing of aerosol samples collected on a filter paper. Near real-time measurements of $PM_{2.5}$ can be achieved by adopting measurement techniques such as optical scattering, beta attenuation method, and tapered element oscillating microbalance (TEOM). Among these, beta attenuation monitors (BAMs) and TEOMs were designated as Federal Equivalent Methods (FEMs) by the USEPA for measuring the hourly mean $PM_{2.5}$. These instruments are expensive and need expertise and infrastructure to install and maintain.

Developed (and high-income) countries have established dense networks of ambient $PM_{2.5}$ monitoring stations. LMICs have spatially sparse $PM_{2.5}$ measurements, mostly confined to urban areas. Across the globe, only 24 countries have at least three $PM_{2.5}$ monitors per million population, while 141 countries have no regular $PM_{2.5}$ monitoring programmes (Martin et al., 2019). The non-uniform geographical distribution of measurements makes it difficult to accurately estimate the health impacts of $PM_{2.5}$ on a global scale. Alternative low-cost methods to improve the spatial coverage of $PM_{2.5}$ include modelling exercises, dense low-cost $PM_{2.5}$ sensor networks, and satellite-based gridded $PM_{2.5}$ estimates.

The Global Burden of Disease (GBD) estimates of air pollution disease and death burden are based on combined $PM_{2.5}$ data from ground-based measurements, global models, and satellite measurements. As polar orbiting satellites provide near-daily global coverage of aerosol measurements, satellite-based $PM_{2.5}$ estimates can significantly improve spatial coverage and uniformity. Reliable and high-resolution estimates of $PM_{2.5}$ are highly sought after to precisely understand and quantify adverse health impacts.

1.2 Literature review

According to the literature reviewed, various statistical, artificial intelligence-based, and hybrid models are available to estimate $PM_{2.5}$ from satellite aerosol products. Aerosol optical depth (AOD), a measure of column-integrated aerosol burden, is the best-suited satellite product to estimate $PM_{2.5}$. However, a standard textbook solution is not available for estimating $PM_{2.5}$ using AOD measurements. An analytical equation relating AOD and $PM_{2.5}$ is available but with a set of assumptions. While AOD is an optical and unit-less column-integrated product, $PM_{2.5}$ is a physical and surface parameter with units ($\mu g m^{-3}$). Also, model performances vary when

adopted across different geographical regions. These aspects point towards the need for region-specific models to estimate $PM_{2.5}$.

Statistical methods include developing simple to advanced models such as (i) simple two-variable linear and non-linear regression models (e.g., Wang & Christopher, 2003; Gupta et al., 2006); (ii) multivariate regression models (e.g., Sreekanth et al., 2017); (iii) linear mixed-effects (LME) models (e.g., Lee et al., 2011; Mhawish et al., 2020); (iv) land-use regression (LUR) models (e.g., Sanchez et al., 2018); (v) geographical weighted regression (GWR) models (e.g., Song et al., 2014; Guo et al., 2017); and (vi) generalised additive models (GAM; e.g., Meng et al., 2018). Some studies developed two-stage/hybrid models for better accuracy in estimating $PM_{2.5}$ (e.g., Dey et al., 2012; Hu et al., 2014).

Artificial intelligence-based methods include neural network (NN) and random forest (RF) models (e.g., Gupta & Christopher, 2009; Mhawish et al., 2020). The scaling of satellite AOD by the factors based on the Chemical Transport Model (CTM) is an extensively used method to estimate national scale-gridded $PM_{2.5}$ (e.g., van Donkelaar et al., 2015; Dey et al., 2012). A recent study by Mandal et al. (2020) used a multistage model (based on ensemble averaging principles) to retrospectively assess the high-resolution daily average $PM_{2.5}$ over the India capital Delhi. Dey et al. (2020) estimated high-resolution $PM_{2.5}$ from satellite AOD using a dynamical scaling factor derived from Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2) reanalysis data.

It is also understood that the performance of the model is based on the choice of auxiliary parameters (predictors) in building the model. Several studies have used collocated meteorological and land-use parameters in improving the model performance because of the impact of these variables on the AOD- $PM_{2.5}$ relationship. Some researchers have incorporated aerosol vertical distribution information from the Cloud-Aerosol Lidar and Infrared Pathfinder (CALIPSO) satellite in their statistical models (Li et al., 2018). In addition to the conventional auxiliary variables, Gupta et al. (2006) studied the influence of cloud cover on the AOD- $PM_{2.5}$ relationship.

In a global study by Christopher and Gupta (2020), spatial and seasonal AOD- $PM_{2.5}$ relationships were assessed in terms of slope, intercepts, and correlation coefficients from a two-variable linear regression analysis. A detailed discussion of the merits and limitations of $PM_{2.5}$ estimated using satellite remote sensing AOD has been provided by Hoff and Christopher (2009) and Duncan et al. (2014) in their studies.

1.3 Need for the study

India is one of the rapidly developing middle-income economies in the world. Continuous urbanisation and industrialisation have deteriorated the air quality in the country, contributing a major share to the air pollution disease burden. High levels of $PM_{2.5}$ is responsible for severe health consequences. In 2019, six Indian cities were listed in the top ten world's most polluted cities¹. The annual population-weighted mean ambient $PM_{2.5}$ was around $90 \mu\text{g m}^{-3}$ for the year 2017 (Balakrishnan et al., 2019), and the current Indian population is around 140 crores. Despite these alarming aspects, India's regulatory air pollution monitoring programme has only 500–600 operating $PM_{2.5}$ stations² (including both manual and continuous).

¹ www.igair.com

² <https://app.cpcbcr.com/ccr/#/caaqm-dashboard-all/caaqm-landing/data>

Studies strongly recommend at least one PM_{2.5} monitor per million population (Martin et al., 2019). As India is an LMIC and the expenditure involved in establishing and maintaining as many monitoring stations is high, a relatively low-cost alternative for the mapping of high-resolution PM_{2.5} is needed. A recent study by Brauer et al. (2019) recommended an integrated (combining modelling and a variety of monitoring methods) approach for India, which includes a satellite-based PM_{2.5} estimation to generate high spatial and temporal resolution PM_{2.5} data. Satellite AOD-based PM_{2.5} (and PM₁₀) estimates for the Indian region were available in the literature (e.g., Kumar et al., 2007; Dey et al., 2012; Chitranshi et al., 2015; Mandal et al., 2020; Dey et al., 2020; Mhawish et al., 2020). Most of these studies focused on Delhi and the Indo-Gangetic Plain (IGP) regions.

This study focuses on building an advanced but relatively less computationally complicated statistical model for satellite AOD-based high-resolution gridded PM_{2.5} estimates over select regions in the southern and northern parts of India. These estimates can aid effective air pollution mitigation and management as air pollution over urban areas is known to vary rapidly in space and time (Apte et al., 2017). Results from the study can help strategise the clean air action plans of cities within the study regions. Furthermore, the study results supports the use of this methodology to develop high resolution PM_{2.5} estimates over large areas of India to help the country achieve the National Clean Air Programme (NCAP) goals. NCAP aims at a 20%–30% reduction in PM pollution by 2024 (base year: 2017; Ganguly et al., 2020).

1.4 Challenges

A major challenge in the use of satellite AOD is data gaps. AOD retrieval is not possible over the pixels in which cloud cover is detected. In addition, satellite orbital gaps and sun glint are other factors contributing to the non-availability of satellite AOD. AOD data are mostly unavailable during the monsoon because of overcast skies.

AOD data are available from various satellite sensors, and each AOD product is characterised by its unique uncertainty envelope. Hence a statistical or artificial intelligence model trained by a particular satellite sensor AOD may not be able to predict PM_{2.5} using a different satellite sensor AOD accurately.

Because of the dynamic nature of the air pollution sources and emissions (especially over LMICs), the spatial and temporal generalisability of a model configuration is questionable. Exercises carried out to overcome these challenges and limitations are detailed in the later parts of the report.

2. Objectives of the Study

In this study, we aim to spatially map the high-resolution satellite AOD-based $PM_{2.5}$ estimates over select regions of India and quantify the estimates for air pollution management and policy recommendations. The specific objectives of the study are as follows:

1. Estimate $1\text{ km} \times 1\text{ km}$ gridded $PM_{2.5}$ using satellite aerosol products
2. Quantify small-scale (i.e., over a few kilometres) and medium-scale (urban–peri-urban or peri-urban–rural) spatial gradients in the estimated $PM_{2.5}$ over select regions
3. Identify $PM_{2.5}$ hotspots within the study region

3. Study Areas and Study Period

The study area included (see Figure 1) the urban, peri-urban, and rural regions of Delhi-NCR, Kanpur (Uttar Pradesh), and Bengaluru (Karnataka). Delhi, Kanpur, and Bengaluru have poor air quality and were also categorised as non-attainment cities (MOEFCC, 2019).

Delhi-NCR: The national capital region (NCR) encompasses Delhi and several districts of adjoining states, viz. Haryana, Uttar Pradesh (UP), and Rajasthan. Delhi is the national capital city of India, while other major cities of the NCR include Noida, Faridabad, Ghaziabad, and Gurugram. Delhi is among the top 10 cities for worst air quality and was also declared India's Asthma Capital³. Geographically, the NCR is a part of the Indo-Gangetic Plain (IGP), which is considered one of the global air pollution hotspots. Local (industrial and transportation) and non-local pollution sources (crop residue burning in neighbouring states), extreme weather patterns (cold winters responsible for a collapsed boundary layer and the frequent haze and fog episodes), geographical barriers (unfavourable conditions for effective horizontal dispersion of pollutants due to the towering Himalayas in the north), and so on aid the build-up of air pollution over Delhi-NCR as well as the IGP region. Air pollution over the IGP region has been extensively studied, characterised, and modelled by several national and international research groups (e.g., Mhawish et al., 2020). The NCR covers around 55,000 km² area and has a population of around 46 million (2011 census) and a population density of around 840 per square kilometre.

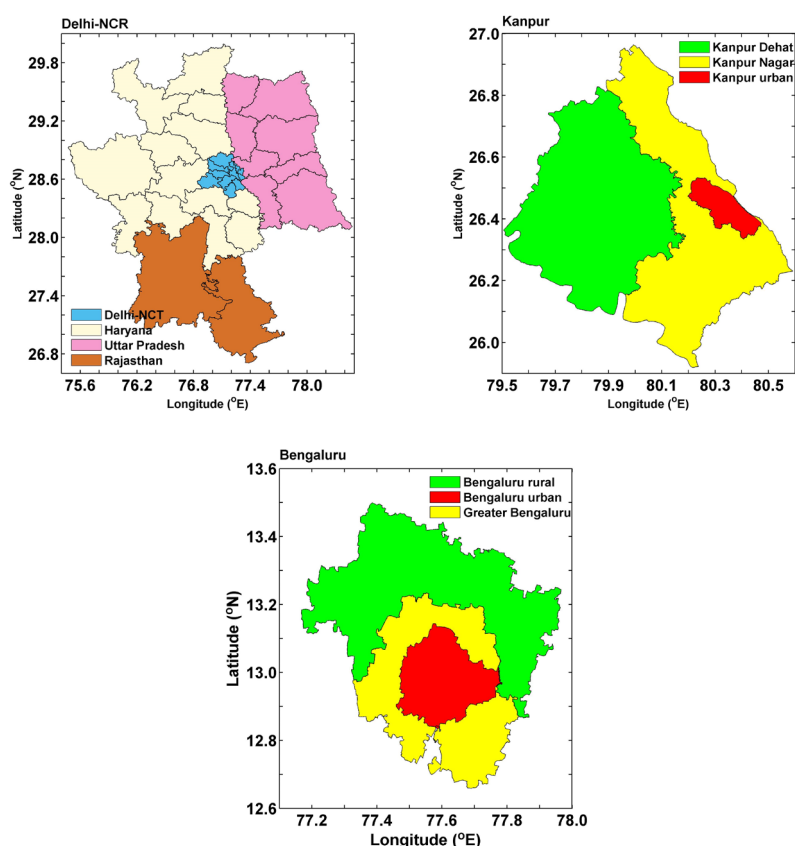


Figure 1: Delhi-NCR (first panel), Kanpur (second panel), and Bengaluru (third panel)

³ <https://urbanemissions.info/delhi-india/>

Kanpur: The Kanpur metropolis is divided into the urban and rural districts of Kanpur, with a total area of around 6,000 km² and a population of 6.4 million (2011 census). The city of Kanpur (located in the Kanpur Nagar district) is the administrative capital of UP and also one of the largest industrial cities in the IGP. Kanpur is the 12th most populous city and the 11th most populous urban agglomeration in India. It is famous for leather and textile industries. More information on Kanpur air pollution, its monitoring status, and management can be found in UrbanEmissions⁴.

Bengaluru: The city of Bengaluru is the administrative capital of the south Indian state of Karnataka. Bengaluru is the third most populous city and the fifth most populous urban agglomeration in India. Bengaluru is considered the silicon valley of India. It houses many information technology (IT) campuses and educational institutions. The metropolitan population is around 11 million, with an area of around 8,005 km². More details on air pollution in Bengaluru can be found in Guttikunda et al. (2019).

Study period

The calendar year 2019 was chosen as the study period. Gridded satellite AOD-based daily mean PM_{2.5} estimation and its spatial mapping were made for the study period. Daily mean PM_{2.5} data were aggregated to monthly, seasonal, and annual scales.

⁴ <https://urbanemissions.info/india-apna/kanpur-india/>

4. Data Sets

4.1 Satellite Aerosol Optical Depth (AOD)

AOD data are available from several satellite sensors. AOD products from Moderate Resolution Imaging Spectroradiometer (MODIS), Multi-angle Imaging SpectroRadiometer (MISR), Ozone Monitoring Instrument (OMI), and Visible Infrared Imaging Radiometer Suite (VIIRS) are some of the most popular and well-evaluated data sets in the world. In this study, we used MODIS AOD (at 0.55 μm), retrieved by the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm. We preferred MODIS-MAIAC-AOD because of its (i) high spatial resolution (1 km x 1 km), (ii) improved retrieval accuracy (compared to its predecessors), and (iii) ability to discriminate between fine and coarse aerosols.

4.1.1 MODIS

MODIS is one of the key sensors aboard the National Aeronautics and Space Administration's (NASA's) Earth Observing System (EOS) satellites: Terra and Aqua. Both satellites orbit in a sun-synchronous, near-polar, circular orbit at an altitude of ~ 705 km. With its swath (cross-track) of $\sim 2,330$ km, MODIS provides global coverage every one or two days. The first MODIS sensor (aboard Terra, equatorial crossing time at 10:30 a.m. LT) was launched in 1999, while the second MODIS sensor (aboard Aqua, equatorial crossing time at 1:30 p.m. LT) was launched in 2002. The passive MODIS sensor measures the Earth-reflected solar radiation in 36 spectral bands ranging between 0.4 and 14.4 μm at varying spatial resolutions (250, 500, and 1000 m). Several geophysical products that can describe the features of land, ocean, and atmosphere can be retrieved from MODIS observations. The MODIS Aerosol Product provides spectral AOD data and aerosol size information⁵.

4.1.2 MAIAC

The satellite AOD retrieval algorithms usually involve (i) the estimation of surface reflectance, (ii) cloud screening, (iii) corrections for gas absorption and, (iv) lookup table (LUT) approach. Conventionally, MODIS-AOD is retrieved using Dark Target (DT) and Deep Blue (DB) algorithms (at a horizontal resolution of ~ 10 km). The DT approach (Levy et al., 2007, 2010) provides relatively accurate estimates of AOD over dark and/or vegetated surfaces. The DB algorithm (Hsu et al., 2004) is developed for and more suitable for AOD retrievals over bright surfaces (e.g., deserts, snow, and arid and urban areas).

The newer MAIAC algorithm, which combines time-series analysis and image-based processing, is aimed at improving the accuracy of cloud detection, aerosol retrievals, and atmospheric correction (Lyapustin, Wang, Laszlo, et al., 2011; Lyapustin et al., 2012, 2018). The MAIAC algorithm uses a physical atmosphere-surface model with minimal assumptions (Lyapustin et al., 2012, 2018) in contrast to the conventional swath-based Lambertian surface model, which is the basis for the DT algorithm. MAIAC retrieves surface Bidirectional Reflectance Distribution Function (BRDF; which defines how light is reflected from an opaque surface) from the collected multi-angle set of images—4 (at poles) to 16 (at the equator) days of observations over each grid.

In addition to the BRDF retrieval, MAIAC characterises unique surface, spectral, spatial, and thermal signatures for each grid cell, which helps improve cloud and snow detection, aerosol

⁵ modis.gsfc.nasa.gov

retrieval, and atmospheric correction (Lyapustin et al., 2018). Then a LUT (pre-computed for a set of aerosol models) approach is adopted for the retrieval of AOD and columnar water vapour (CWV). The MAIAC daily atmosphere product includes 1 km gridded (on a global sinusoidal projection) (i) CWV (based on MODIS near-infrared observations at 0.94 μm), (ii) AOD (at 0.47 and 0.55 μm), (iii) cloud mask, (iv) aerosol type, and (v) smoke injection height. Combined MAIAC data (for both the Terra-MODIS and Aqua-MODIS satellite sensors) are available in the hierarchical data format (HDF4), data divided into 1200 x 1200 km^2 standard MODIS tiles.

The initial evaluation of MAIAC retrievals has shown better sensitivity to small dust storms but exhibited biases over bright desert regions (Lyapustin, Wang, Hsu, et al., 2011). Several key components of the MAIAC algorithm have been significantly improved over the years since its inception. Several recent validation studies across various parts of the globe have demonstrated better retrieval accuracies for MAIAC-AOD compared to VIIRS-AOD, DT-AOD, and DB-AOD (Martins et al., 2017; Superczynski et al., 2017; Jethva et al., 2019; Mhawish et al., 2019). In a recent comparison study over South Asia, Mhawish et al. (2019) found that MAIAC-AOD exhibited lower bias (compared to DB-AOD and DT-AOD) for ground truth Aerosol Robotic Network (AERONET) AOD measurements.

4.2 $\text{PM}_{2.5}$

In this study, ambient $\text{PM}_{2.5}$ data were acquired from the regulatory measurements being carried out by India's Central and State Pollution Control Boards (CPCB/SPCB) under the Continuous Ambient Air Quality Monitoring Stations (CAAQMS) programme. The CAAQMS programme uses reference-grade instruments to measure the 'criteria' air pollutants. In this programme, a beta attenuation monitor (BAM) measures near-real-time (hourly) $\text{PM}_{2.5}$. Technical details of the instrumentation used and measurement protocols can be found elsewhere⁶. BAM measurements are based on the Beer-Lambert law, which states that the amount by which the flow of beta particles is attenuated by particulate matter is exponentially dependent on their mass. A BAM that measures $\text{PM}_{2.5}$ will be equipped with a sharp cut 2.5 μm cyclone (to remove larger particles from the airflow entering the measurement chamber), heater (to remove moisture from the ambient airflow), pump (to draw the ambient air at a prescribed flow rate), radioactive source (to emit beta particles), a scintillation detector (to measure the transmitted beta particles), and a filter tape (to collect the PM sample).

4.3 Meteorological data

Meteorological variables used in the study include surface (2 m) air temperature (ST , K), surface (2 m) relative humidity (RH , %; derived from 2 m dew point temperature), wind fields (10 m; wind direction (WD , °) and wind speed (WS , m s^{-1}); derived from separate u and v wind components), planetary boundary layer height ($PBLH$, m), and surface (2 m) pressure (SP , Pa). Gridded hourly data for these meteorological variables were acquired from the European Centre for Medium-Range Weather Forecast's (ECMWF's) reanalysis database. ECMWF combines its forecast models and data assimilation systems to produce reanalysed global data sets that are complete and consistent. Among several versions of reanalysis data sets, ECMWF Reanalysis-5 Land (ERA5-Land; data available from 1981 to present) derived meteorological data (except $PBLH$) were used. ERA5-Land is produced by replaying the land component of

⁶ <http://www.ppcb.gov.in/Attachments/Tenders/Technical.pdf>

ERA5 climate reanalysis. The native horizontal resolution of ERA5-Land⁷ is around 9 km. PBLH data were acquired from the ERA5 product⁸. The horizontal grid size of ERA5 is around 30 km.

4.4. Auxiliary data

Auxiliary predictors used in the study include Normalized Difference Vegetation Index (NDVI) and Elevation (ELV, m). NDVI was considered a proxy for land use. Level-3 gridded 1 km MODIS-derived NDVI data were used. MODIS NDVI is produced at 16 days temporal interval at different horizontal resolutions. The NDVI retrieval is based on the atmosphere-corrected bidirectional surface reflectance. NDVI is a normalised transform of the NIR to red reflectance ratio, which varies between -1 and +1. Leveraging the identical MODIS sensors aboard Terra and Aqua satellites, the MODIS vegetation index (VI) algorithm generates each sensor's 16-day composite eight days apart. This enables a higher temporal resolution of the NDVI product.

ELV data are acquired from the global digital elevation model (DEM), ETOPO2 of the National Oceanic and Atmospheric Administration (NOAA). ETOPO2 is a 2 arc-minute global DEM, which includes both land topography and ocean bathymetry. The horizontal resolution⁹ of ETOPO2 is around 3.7 km. A summary of the predictors is tabulated below.

Table 1: The list of shortlisted predictors used for model building.

Predictor	Data source	Spatial resolution	Temporal resolution
Aerosol optical depth (AOD)	MODIS satellite sensor (aboard Terra and Aqua)	1 km	Daily
Columnar water vapour (CWV)	MODIS satellite sensor (aboard Terra and Aqua)	1 km	Daily
Normalised Difference Vegetation Index (NDVI)	MODIS satellite sensor (aboard Terra and Aqua)	1 km	8-day
2 m temperature (ST)	ERA5-Land	9 km	Hourly
Relative humidity	ERA5-Land (derived from temperature and dew point temperature)	9 km	Hourly
Wind fields (speed and direction)	ERA5-Land (derived from u and v components)	9 km	Hourly
Planetary boundary layer height (PBLH)	ERA5	30 km	Hourly
Surface pressure	ERA5-Land	9 km	Hourly
Elevation	ETOPO2	3.7 km	NA

⁷ <https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5-land>

⁸ <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>

⁹ More details on ETOPO2 can be found here: <https://www.ngdc.noaa.gov/mgg/global/etopo2.html>



5. Data Handling

5.1 AOD

All the MODIS products were downloaded from the NASA website¹⁰. Daily files for the Terra-Aqua MAIAC AOD product (MCD19A2) were downloaded for the study regions (respective tiles). Only the highest quality AOD data with the quality assurance (QA) flag 'clear' were retained for further processing. By leveraging the daily twin MODIS-AOD measurements (Terra and Aqua), we imputed the missing AOD values to improve their spatial coverage. On a daily basis, we established a linear relationship between the simultaneous Terra- and Aqua-AOD over the study region, and the obtained regression coefficients were used to estimate Terra-AOD when only Aqua-AOD was available and vice versa (Huang et al., 2018). Thresholds on the linear regression coefficients (slope, intercept, and R^2) were imposed to achieve reliable imputations. Unreliable regression coefficients could be because of the small number of Terra- and Aqua-AOD matchups on that day (because of several known reasons). The efficiency of the imputation method in increasing the spatial coverage of AOD is shown in Figure A1 for a typical winter day. Data gaps (over Karnataka) can be seen in both Terra- and Aqua-AODs because of the satellite orbital gap (the top row of the figure). These gaps are filled (the bottom row of Figure A1) after the imputation process. In this study, the imputation process was carried out separately for study regions in southern and northern India. After this process, grid-wise daily Terra- and Aqua-AODs were averaged to represent the daily mean AOD.

5.2 PM_{2.5}

Hourly PM_{2.5} data were downloaded from the CPCB dashboard¹¹. Hourly values were averaged to compute the daily mean PM_{2.5}. Data quality check was performed on the hourly PM_{2.5} before averaging. This included (i) the removal of unrealistic PM_{2.5} values (instrument fill values, large negatives), (ii) the removal of outliers (hourly PM_{2.5} values outside the range of the daily mean $\pm 3 \times$ standard deviation on a log-transformed scale were flagged as outliers), (iii) the removal of all PM_{2.5} values greater than PM₁₀ (as PM_{2.5} is a subset of PM₁₀), and (iv) a 75% completeness criteria for daily averaging (at least 18 hours data in a day must be available to compute the daily mean). In Figure A2, a daily PM_{2.5} time series is shown before and after the data cleaning process. It can be observed that spurious peaks in the data are not present after processing for outliers and unrealistic values.

5.3 Meteorological parameters

Hourly mean meteorological variables were downloaded from the climate data store of the European Commission¹².

5.3.1 Derived variables

ERA5-Land provides u (zonal velocity) and v (meridional velocity) components of the wind field separately. From these two components, WS and WD are derived using the following equations:

$$WS = \sqrt{u^2 + v^2} \quad \dots\dots\dots (1)$$

¹⁰ <https://ladsweb.modaps.eosdis.nasa.gov/>

¹¹ <https://app.cpcbcr.com/ccr/#/caaqm-dashboard-all/caaqm-landing>

¹² <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview>

$$WD = \text{atan2}(v, u) \quad \dots\dots\dots (2)$$

ERA5-Land provides surface (2 m) dew point temperature (DPT), from which RH was derived using the equation below following Alduchov and Eskridge (1996). ST and DPT need to be converted into degrees Celsius (°C) before applying Equation 3.

$$RH = 100 * \frac{\exp\left(\frac{17.625 * DT}{243.04 + DT}\right)}{\exp\left(\frac{17.625 * ST}{243.04 + ST}\right)} \quad \dots\dots\dots (3)$$

5.3.2 Spatial interpolation

Interpolation techniques were used to match up the grid size of ERA-5 meteorological parameters and ETOPO2 elevation data to the MODIS-MAIAC grid size. The 9 km ERA5-Land and 30 km ERA5 daily mean meteorological variables are spatially interpolated to the 1 km MAIAC grid using the bilinear interpolation method following Mhawish et al. (2020). The same technique was used to interpolate the 3.7 km elevation data. Figure A3 shows ST and PBLH data (over the Greater Bengaluru area) before and after spatial interpolation for a typical winter day.

5.4 NDVI

Both Terra- and Aqua-MODIS L3 NDVI products (MOD13A2 and MYD13A2) were downloaded for the study region and study period. Each satellite sensor provides 16-day composites. Combining NDVI data from both sensors, 8-day temporal resolution can be achieved. Daily NDVI values (over each 1 km grid) were estimated by temporally interpolating the observations. A cubic spline technique was used for interpolation following Mhawish et al. (2020). Figure A4 showcases the NDVI observations and temporally interpolated values over a typical 1 km spatial grid for a year.

6. Methodology

6.1 Model selection

Being a developing country, the $PM_{2.5}$ load in India and its sources can be highly variable in time and space. These, in turn, can be responsible for the spatial and temporal heterogeneity in the aerosol optical and microphysical properties and their vertical distribution. These aspects suggest a model that can calibrate AOD (to estimate $PM_{2.5}$) at a high resolution both spatially and temporally. An LME model is one such statistical approach that can accommodate fixed and random effects. The LME model is highly suitable for grouped data and also when groups have an unequal number of observations. The LME model has proven to be effective (cross-validation [CV] R^2 greater than 0.8) in estimating satellite AOD-based $PM_{2.5}$ over low (e.g., New England region) and highly polluted (e.g., Beijing) environments (Lee et al., 2011; Xie et al., 2015; Han et al., 2018; Mhawish et al., 2020). Based on the daily grouping of 10 km satellite AOD and ground $PM_{2.5}$, Lee et al. (2011) observed the CV R^2 greater than 0.92 over the New England region. In a sensitivity exercise, Xie et al. (2015) demonstrated that the use of finer-resolution AOD in the LME model can yield better model performance. Over Beijing, Han et al. (2018) observed a superior LME model performance (CV R^2 greater than 0.92) in predicting $PM_{2.5}$ from 1 km MAIAC-AOD. Mhawish et al. (2020) configured their LME model with day-specific and month-specific random effects in the AOD- $PM_{2.5}$ relationship and also added meteorological parameters (as fixed effects) to predict $PM_{2.5}$ over the IGP. Based on the understanding obtained from these studies, the configuration of the LME model that was adopted in the current study is as follows:

$$PM_{2.5\ i,j} = (\alpha_o + \alpha_{day}) + (\beta_o + \beta_{day}) \times AOD_{i,j} + \beta_1 \times CWV_{i,j} + \beta_2 \times ST_{i,j} + \beta_3 \times RH_{i,j} + \beta_4 \times WD_{i,j} + \beta_5 \times WS_{i,j} + \beta_6 \times PBLH_{i,j} + \beta_7 \times SP_{i,j} + \beta_8 \times NDVI_{i,j} + \beta_9 \times ELV_{i,j} + \varepsilon_{i,j} \dots (4)$$

i corresponds to the monitoring site for $PM_{2.5}$, while for the other variables, it represents the nearest (to the corresponding monitoring site) grid. j corresponds to the day. α_{day} and β_{day} correspond to the day-specific random intercept and random slope, while other regression coefficients are fixed effects. $\varepsilon_{i,j}$ is the error term.

As $PM_{2.5}$ estimates obtained in the study are intended to be used for formulating policy recommendations, the model is configured to predict the daily mean $PM_{2.5}$ (for which a national threshold of $60 \mu\text{g m}^{-3}$ is prescribed). Hereafter, $PM_{2.5}$ represents the daily mean values. In the case of AOD, Terra-MODIS and Aqua-MODIS average were considered the daily mean. MODIS NDVI is only one value per day (after temporal interpolation), and ELV is a spatial variable and temporally invariant.

In a sensitivity exercise, we found that the $PM_{2.5}$ predictions were better (in terms of accuracy and precision) (i) when meteorological and auxiliary variables were also considered predictors (in addition to AOD) and (ii) when configured with daily mean predictors. We also observed no difference in the model performance with and without WD and CWV inclusion.

Equation 4 assumes that the day-to-day variability in the AOD- $PM_{2.5}$ linear relationship is spatially invariant. Using data from 35 $PM_{2.5}$ monitoring stations (which spread across the Beijing municipality area of $\sim 16,400 \text{ km}^2$), Xie et al. (2015) have demonstrated marginal improvement in the model performance when site-specific random effects were introduced (in addition to day-specific random effects). Also, the incorporation of site-specific random effects hinders predicting $PM_{2.5}$ over grids with no monitoring sites (unless the site-specific random regression coefficients are interpolated over the study region). Site-specific random effects

were not incorporated in the current model; instead, two separate models were configured each for the North Indian (Delhi-NCR and Kanpur) and South Indian (Bengaluru) study regions.

6.2 Random forest (RF) model

Random Forest (a supervised machine learning algorithm) is an ensemble method that uses results from multiple decision-tree regressors to decrease the variance of the developed model. In this method, multiple decision trees are modelled using samples drawn from the data set without replacement, also known as the bagging approach. The results from the decision trees are then averaged to give the final result of the random forest. In the case of classification, the result from the most number of decision trees is used as the final result. In the current study, the RF model (to estimate the daily mean PM_{2.5}) was developed using the random forest regressor module of the scikit-learn library (Python).

6.3 Model validation

To assess the precision and accuracy of the model predictions, we performed a 10-fold CV and leave-one-out (LOO) CV. CV is a statistical method to investigate the model performance on unseen data. In the 10-fold CV exercise, the total data were randomly split into 10 equal groups. The model is trained with 9 data groups (training groups), and the data in the tenth group (test group) were predicted and compared. The process was repeated 10 times until each unique group was used as the test group. In the LOOCV exercise, the model is trained with data from all stations except one particular station. Data from the remaining station were predicted and compared. The process was repeated k (number of stations) times until each unique station data were predicted.

The precision and accuracy of the model predictions were quantified using the following statistical parameters:

1. Coefficient of determination (R²)
2. Mean absolute bias (MAB)
3. Root Mean Square Error (RMSE)
4. Normalised Root Mean Square Error (NRMSE)

$$R^2 = \left(\frac{\sum_{i=1}^n P_i O_i - \frac{1}{n} \sum_{i=1}^n P_i \sum_{i=1}^n O_i}{\sqrt{\sum_{i=1}^n P_i^2 - \frac{1}{n} (\sum_{i=1}^n P_i)^2} \sqrt{\sum_{i=1}^n O_i^2 - \frac{1}{n} (\sum_{i=1}^n O_i)^2}} \right)^2 \dots\dots\dots (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \dots\dots\dots (6)$$

$$MAB = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \dots\dots\dots (7)$$

$$NRMSE = \frac{RMSE}{\frac{1}{n} \sum_{i=1}^n O_i} \dots\dots\dots (8)$$

where *P* and *O* are the predictions and observations, respectively, and *n* is the number of data pairs.

6.4 Monitoring stations

We used hourly PM_{2.5} data from 89 CAAQMS (Figure 2) spread across Delhi-NCR, UP, and Karnataka. Data from all these urban stations were considered to build the LME models. Out of

the 89 stations, 14 were from Karnataka and the rest were from Delhi-NCR and UP. The list of CAAQMS along with the geographical coordinates and yearly mean (2019) PM_{2.5} and AOD over those station locations are given in Table A1 (see Appendix A).

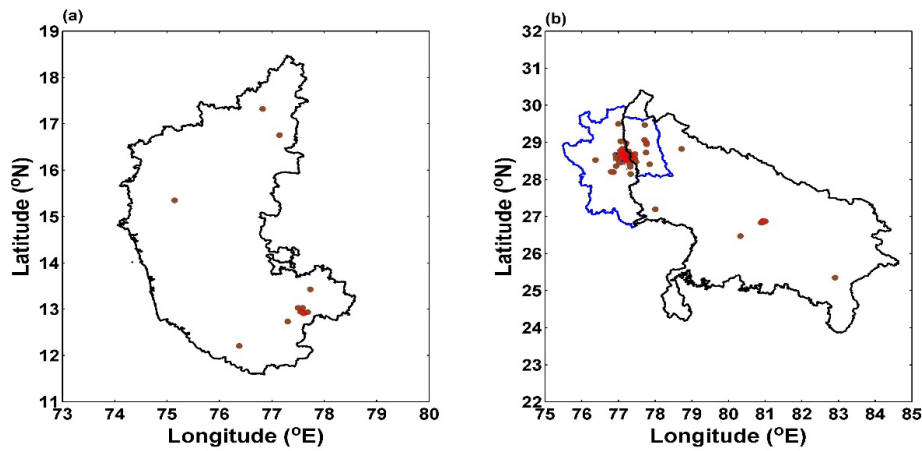


Figure 2: Geographical locations of the 89 CAAQMS overlaid on the state boundaries of (a) Karnataka and (b) UP. The blue line in panel (b) represents Delhi-NCR.

6.5 Hotspot analysis

Hotspots are relatively small areas (or pixel clusters) characterised by higher attribute values in comparison with their surrounding areas. In this study, we used statistical techniques based on spatial autocorrelation to identify PM_{2.5} hotspots over the study regions. We used an ArcGIS-based hotspot analysis tool that calculates the Getis-Ord Gi* statistic (Getis & Ord, 1992) to examine the spatial patterns of PM_{2.5} at a local scale. The application of this tool returns z-scores (Gi* statistic) and p-values that help in identifying the spatial clusters of high (hotspots) and low (cold spots) values and their statistical significance. To qualify as a statistically significant hotspot, an identified feature with high value should be surrounded by other features with high values—vice versa for a cold spot. A statistically significant higher positive Gi* statistic indicates an intense clustering of high values. Similarly, a statistically significant smaller negative Gi* statistic indicates intense clustering of low values. This methodology has been widely used to identify pollution hotspots (Habibi et al., 2017). Gi* statistic is calculated using the following equations:

$$Gi^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n (w_{i,j})^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}} \dots\dots\dots (9)$$

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \dots\dots\dots (10)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \dots\dots\dots (11)$$

In the above equations, x_j is the attribute value (PM_{2.5} in this study) for feature j , n is the total number of features, and $w_{i,j}$ is the spatial weights between features i and j .

6.6 Rural, peri-urban, urban, and uninhabited pixel classification

To identify the land cover type, Global Human Settlement Layer Settlement Model (GHSL-SMOD) data were used. GHSL utilises data from high-resolution satellite imagery, census, and volunteered geographic information. GHSL-SMOD identifies human settlements and their extent by mapping the types of urban areas consistently and systematically across the globe

(Melchiorri et al., 2018; Chowdhury et al., 2019). In the present study, the 1 km pixels were categorised into 4 classes: uninhabited (very low-density settlement), rural (small and low-density settlement), peri-urban (semi-settlement), and urban (large, dense, and medium settlement).

7. Results

7.1 Spatio-temporal variations in the daily mean PM_{2.5} and AOD

Box plots shown in Figure 3 depict spatio-temporal variations in the daily mean (derived from hourly CAAQMS measurements) PM_{2.5} for the year 2019. The calendar year is classified into four seasons, viz. winter (WIN; January, February), summer (SUM; March, April, May), monsoon (MON; June, July, August, September), and post-monsoon (PMN; October, November, December). In the box plots, the solid dot represents the mean, the central line of the box represents the median, the boxes represent the 25 and 75 percentiles, and whiskers represent the minimum and maximum values of the distribution.

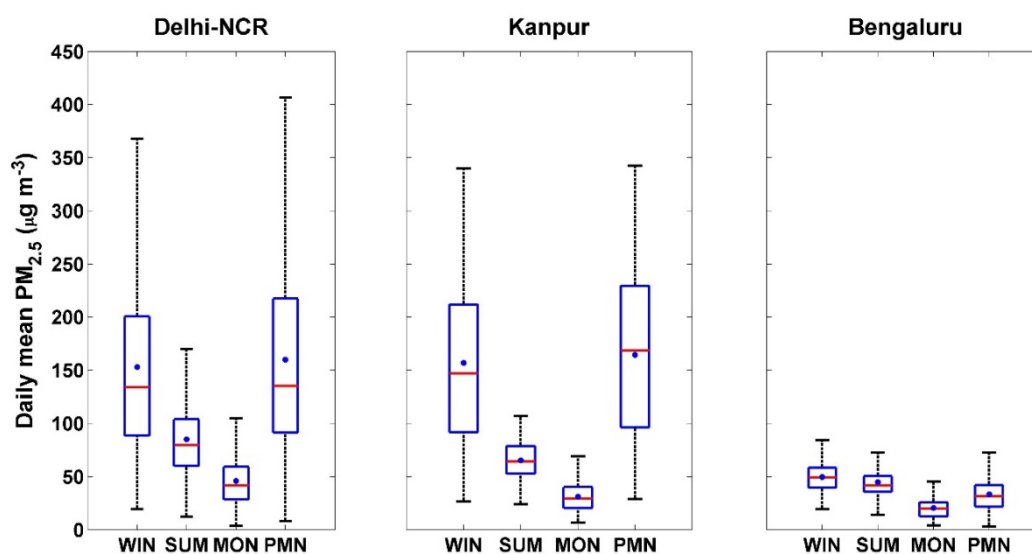


Figure 3: Spatio-temporal variations in the daily mean PM_{2.5}. Individual stations' data from each region were combined.

In line with the earlier studies, the CAAQMS-measured PM_{2.5} values over the northern Indian study regions (Delhi-NCR and Kanpur) were much higher than the values observed over the southern Indian study region (Bengaluru). PM_{2.5} over Delhi-NCR and Kanpur were comparable. The seasonal mean PM_{2.5} over Delhi-NCR and Kanpur ranged between 165 ± 77.6 and 31 ± 14.6 $\mu\text{g m}^{-3}$. Over Bengaluru, it ranged between 50 ± 13.6 and 21 ± 9.9 $\mu\text{g m}^{-3}$. Seasonally, the highest mean values were observed during PMN/WIN and the lowest during MON across the study regions. Seasonal PM_{2.5} statistics are shown in Table 2.

Variations in the daily mean AOD (Terra-MODIS and Aqua-MODIS mean AOD) over the CAAQMS locations (nearest grid) are shown in Figure 4. The absolute values of AOD and their seasonal spread are higher over Delhi-NCR and Kanpur. Heterogeneity in the seasonal AOD pattern is observed across the study regions. Over Bengaluru, SUM AOD is higher than WIN AOD, while a contrasting pattern can be observed over Delhi-NCR and Kanpur. The observed Bengaluru seasonal AOD variations followed the climatological pattern shown in the study by Sreekanth (2013). Seasonal high AOD values were observed during PMN over Delhi-NCR and Kanpur, while it was during SUM over Bengaluru. Seasonal AOD statistics are shown in Table 3.

Seasonal box plots for all other predictors are shown in Appendix A (Figures A5–A10).

Table 2: Season-wise daily mean PM_{2.5} statistics. IQR and SD represent the interquartile range and standard deviation, respectively.

Study region	Season	Minimum ($\mu\text{g m}^{-3}$)	Maximum ($\mu\text{g m}^{-3}$)	Mean ($\mu\text{g m}^{-3}$)	SD ($\mu\text{g m}^{-3}$)	Median ($\mu\text{g m}^{-3}$)	IQR ($\mu\text{g m}^{-3}$)
Delhi-NCR	WIN	19	477	153	84.1	134	112.3
	SUM	12	302	85	35.3	80	44.0
	MON	4	273	46	24.4	41	30.9
	PMN	8	492	160	93.3	135	126.4
Kanpur	WIN	27	340	157	80.7	147	120.2
	SUM	24	107	65	18.3	64	25.8
	MON	7	78	31	14.6	29	20.1
	PMN	29	343	165	77.6	169	133.2
Bengaluru	WIN	19	99	50	13.6	49	18.7
	SUM	14	124	45	16.1	42	14.7
	MON	4	93	21	9.9	20	13.0
	PMN	3	110	33	16.4	31	20.5

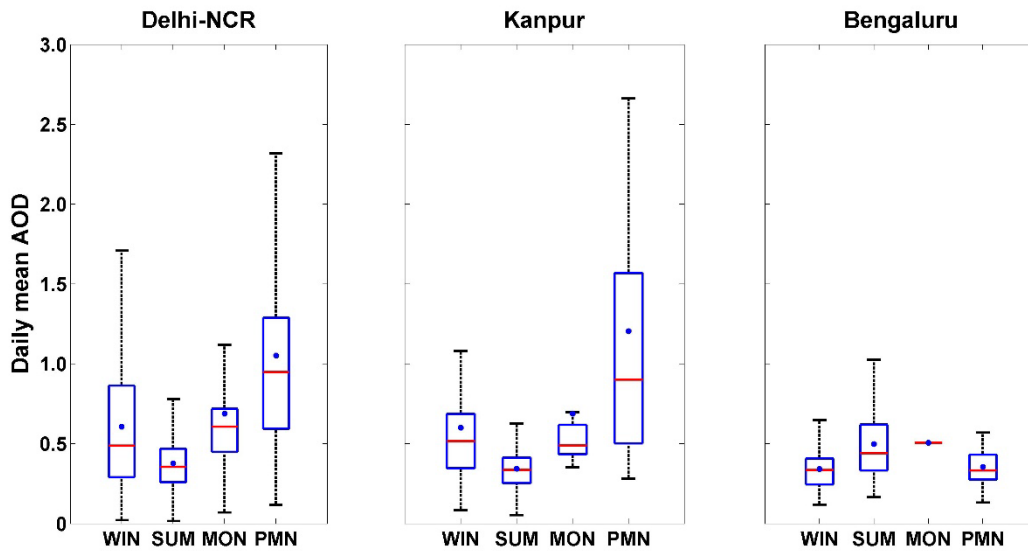


Figure 4: Spatio-temporal variations in the daily mean AOD

Table 3: Season-wise AOD statistics. IQR and SD represent the interquartile range and standard deviation, respectively.

Study region	Season	Minimum	Maximum	Mean	SD	Median	IQR
Delhi-NCR	WIN	0.021	2.609	0.608	0.411	0.487	0.575
	SUM	0.015	1.536	0.378	0.175	0.356	0.208
	MON	0.015	2.879	0.688	0.442	0.607	0.270
	PMN	0.117	3.276	1.053	0.648	0.947	0.697
Kanpur	WIN	0.084	2.738	0.601	0.461	0.515	0.342
	SUM	0.054	0.814	0.343	0.126	0.337	0.161
	MON	0.353	2.337	0.689	0.541	0.489	0.184
	PMN	0.281	3.276	1.206	0.939	0.902	1.067
Bengaluru	WIN	0.119	0.817	0.341	0.134	0.334	0.162
	SUM	0.165	1.097	0.500	0.212	0.441	0.286
	MON	0.507	0.507	0.507	0.000	0.507	0.000
	PMN	0.132	0.963	0.356	0.125	0.333	0.157

7.2 Model evaluation

Figures 5 and 6 show the model (Equation 5) performance in terms of 10-fold CV and LOOCV. Figure 5 corresponds to Delhi-NCR and Kanpur model, while Figure 6 is for the Bengaluru model. In Figures 5 and 6, the black solid line represents the 1:1 line, while the red solid line is the linear regression fit to the data. Statistics are given in each of the panels.

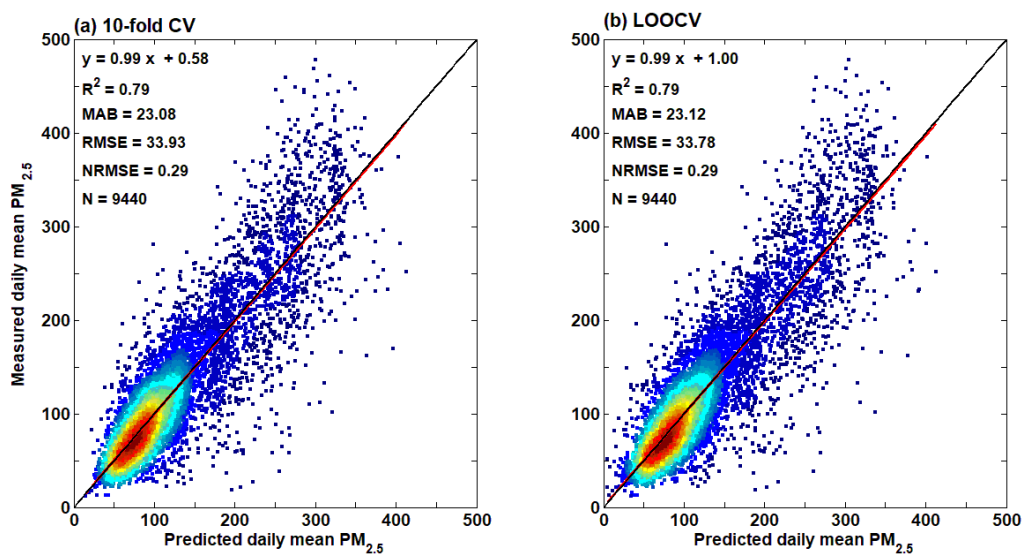


Figure 5: Density scatter plots depicting the model-predicted daily mean $PM_{2.5}$ and measured daily mean $PM_{2.5}$ (the Delhi-NCR and Kanpur model)

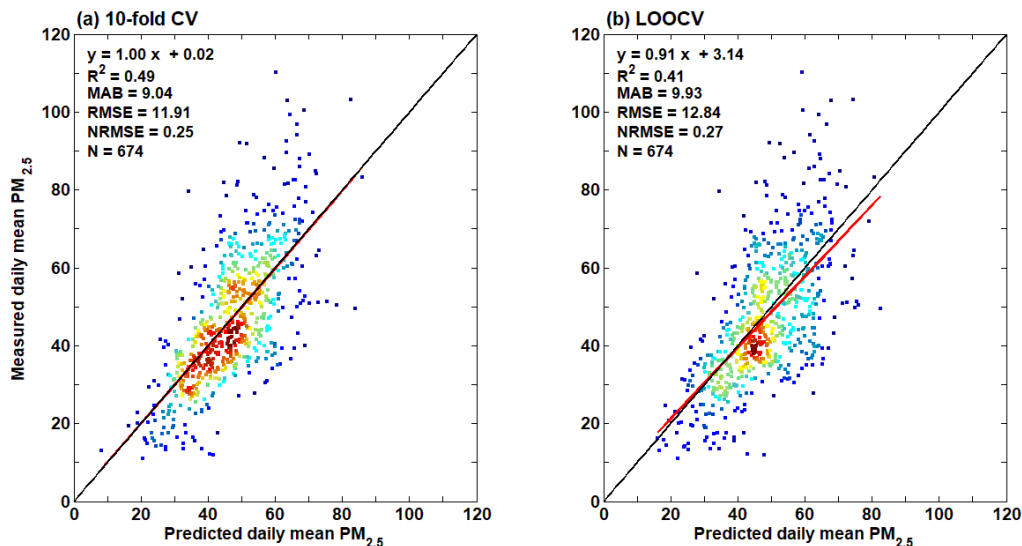


Figure 6: Density scatter plots depicting the model-predicted daily mean $PM_{2.5}$ and measured daily mean $PM_{2.5}$ (the Bengaluru model)

The Bengaluru model was less efficient in terms of the coefficient of determination (R^2). This could be because of the availability of fewer data points for model training. Also, R^2 values are highly sensitive to outliers. But the normalized root mean square error (NRMSE) values are comparable between both the models. In both models, the R^2 value decreased marginally during cross-validation. To evaluate the model performance at different temporal averaging scales, the monthly mean from the predicted daily $PM_{2.5}$ and the monthly mean from the measured $PM_{2.5}$ were compared and are shown in Figure A11. LME model predictions were more accurate when daily averages were aggregated to the monthly scale. All the performance statistics were improved for the monthly comparison. For the Bengaluru model, the 10-fold CV R^2 improved from 0.49 to 0.63 (Figure A11). The seasonal scale (plots not shown because of the small number of data points) averaging has further improved the performance statistics. A majority of the results obtained in the present study were presented on seasonal scales.

For comparison, we also trained another LME model, incorporating site-specific random effects (random intercept) in addition to day-specific random effects. Figure A12 corresponds to Delhi-NCR and Kanpur model, while Figure A13 is for the Bengaluru model. We observed a marginal improvement in site- and day-specific random-effects model performance when compared to the performance of the day-specific only random-effects model. We haven't adopted the site-specific and day-specific random effects LME model in the current study as the site-specific random coefficients need to be interpolated for the grids in which no monitoring station is located.

In addition to statistical models, we also constructed a machine learning-based model to compare its performance with that of the LME model. The performance (in terms of CV) of the random forest (RF) model (for Delhi-NCR and Kanpur) is shown in Figure A14. In terms of R^2 , MAB, RMSE, and NRMSE, both LME and RF models performed almost equally. These results are in contrast to the observations of Mhawish et al. (2020). This could be because the estimate in the current study is the daily mean $PM_{2.5}$, while it is 10 to 14 h LT mean $PM_{2.5}$ in Mhawish et al. (2020). Based on these results, we adopted the LME (with day-specific random effects) model for spatial predictions of the daily mean $PM_{2.5}$.

All the training data used in the current study are urban-centric. The LME model trained with urban $PM_{2.5}$ was used for regional spatial predictions, which include rural and peri-urban areas.

To investigate the model's efficacy in predicting rural PM_{2.5}, BAM-measured PM_{2.5} data from a rural location in UP were considered. The BAM is being operated by the Indo Gangetic Plains Center for Air Research and Education (IGPCARE) research station in the Ruri Para village (25.815°N, 79.918°E), Hamirpur district of UP. The LME (Delhi-NCR and Kanpur model) predicted daily mean PM_{2.5} and the BAM measured PM_{2.5} were compared (Figure A15) for the SUM season. Grossly, the models' predictions overestimate with an NRMSE value of 0.22. This value is less compared to CV NRMSE (Figure 5), indicating the comparable performance of the model in predicting urban and rural PM_{2.5}.

7.3 Spatial maps of PM_{2.5}

7.3.1 Delhi-NCR region

Figure 7 shows the annual mean map of PM_{2.5} over Delhi-NCR. Over each 1 km grid, individual predicted daily mean PM_{2.5} values for the year 2019 were averaged to obtain the annual mean PM_{2.5}. The spatially averaged annual mean PM_{2.5} for the year 2019 for Delhi-NCR was $101 \pm 7.5 \mu\text{g m}^{-3}$ (value after \pm is the standard deviation). Grid-wise annual mean PM_{2.5} values ranged between 80 and $130 \mu\text{g m}^{-3}$. These values could be an overestimate owing to the data gaps in the estimated MON PM_{2.5}.

During MON, PM_{2.5} levels are expected to be lower than normal. This data gap can lead to an overestimation of the annual mean PM_{2.5} over a particular grid. Using the CAAQMS-measured and the corresponding grid-LME-predicted daily mean PM_{2.5}, the overestimation in the model predicted annual mean values was found to be $\sim 12\%$. The annual mean PM_{2.5} was found to exceed the Indian annual threshold ($40 \mu\text{g m}^{-3}$) over all the pixels of Delhi-NCR. The highest PM_{2.5} levels ($\sim 130 \mu\text{g m}^{-3}$) were observed over the National Capital Territory (NCT) region.

The districts of Gurugram, Faridabad, Nuh, Palwal, and Bharatpur in the NCR also observed higher concentrations of PM_{2.5} in the range $105\text{--}120 \mu\text{g m}^{-3}$. Over other regions of the NCR, the PM_{2.5} range was observed to be $80\text{--}100 \mu\text{g m}^{-3}$. Figure 7(b) shows the frequency distribution of gridded annual mean PM_{2.5} values over Delhi-NCR for the study year. There were a total of 64,221 grids (1 km size) covering Delhi-NCR. The frequency distribution of PM_{2.5} was right-skewed, and $\sim 86\%$ of the PM_{2.5} values were observed in the range of $90\text{--}110 \mu\text{g m}^{-3}$. About 2% of the data exceeded $120 \mu\text{g m}^{-3}$ (three times the annual standard).

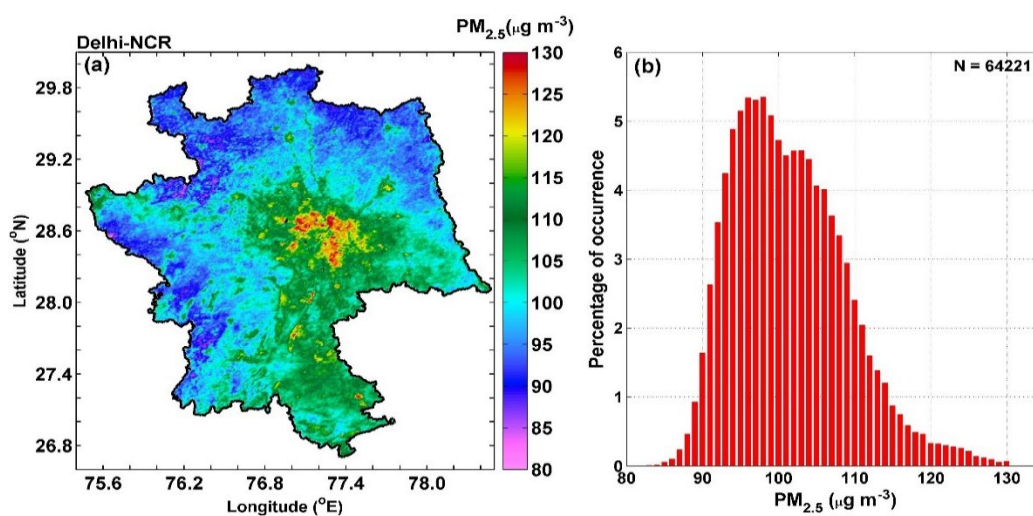


Figure 7: The spatial map of the annual mean PM_{2.5} over Delhi-NCR and its frequency distribution

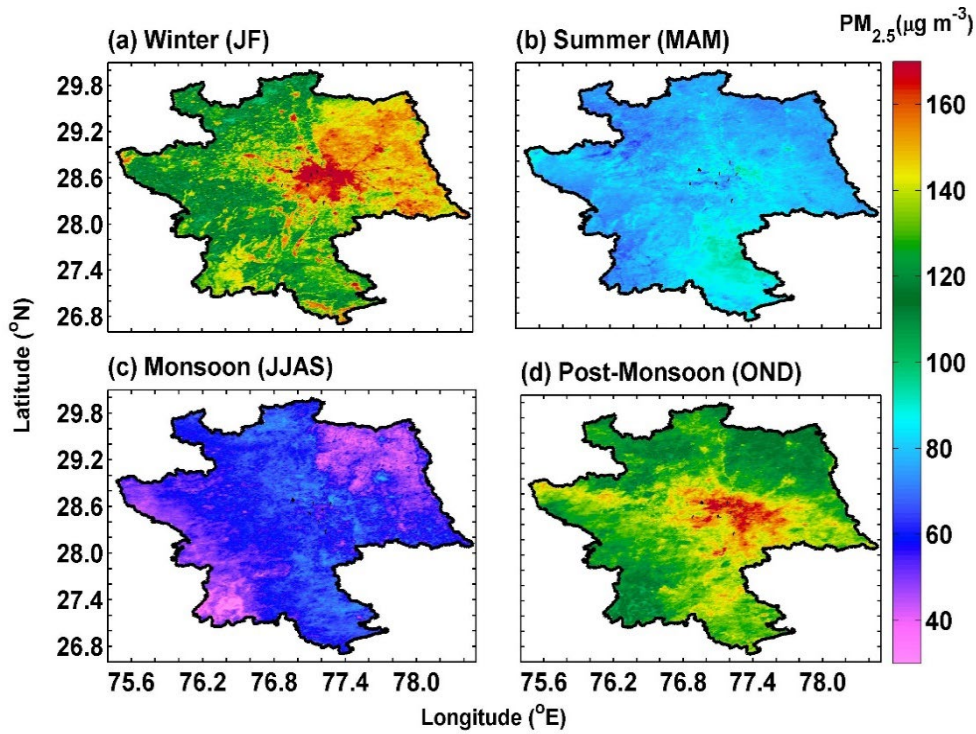


Figure 8: The seasonal mean $PM_{2.5}$ maps over Delhi-NCR

Figure 8 shows seasonal mean maps of $PM_{2.5}$ over Delhi-NCR. The seasonal maps exhibited significant spatial and temporal variations in $PM_{2.5}$ across Delhi-NCR. Seasonally, WIN and PMN were the most polluted seasons, while MON was characterised by the lowest $PM_{2.5}$ levels. The highest $PM_{2.5}$ values were observed during WIN ($134 \pm 16.0 \mu\text{g m}^{-3}$) followed by PMN ($131 \pm 11.9 \mu\text{g m}^{-3}$), SUM ($79 \pm 5.3 \mu\text{g m}^{-3}$), and MON ($57 \pm 8.2 \mu\text{g m}^{-3}$). See Table 4 for more statistics. Over the NCT region, WIN and PMN $PM_{2.5}$ were as high as $\sim 160 \mu\text{g m}^{-3}$. Higher values were also identified in the eastern districts of NCR during WIN.

7.3.2 Kanpur region

The annual mean spatial map of the predicted $PM_{2.5}$ over the Kanpur region is shown in Figure 9. The spatial map includes both rural (Kanpur Dehat) and urban (Kanpur Nagar) districts of Kanpur. The spatially averaged annual mean $PM_{2.5}$ for the year 2019 was $108 \pm 4.5 \mu\text{g m}^{-3}$. Over Kanpur urban, $PM_{2.5}$ values were as high as $130 \mu\text{g m}^{-3}$. The frequency distribution of the gridded spatial $PM_{2.5}$ values is shown in Figure 9(b). There were a total of 7,189 grids (1 km size) covering both the rural and urban areas of Kanpur. The frequency distribution exhibits a log-normal distribution with modal $PM_{2.5}$ values $\sim 108 \mu\text{g m}^{-3}$. The annual mean $PM_{2.5}$ over all the spatial grids exceeded the national annual standard ($40 \mu\text{g m}^{-3}$).

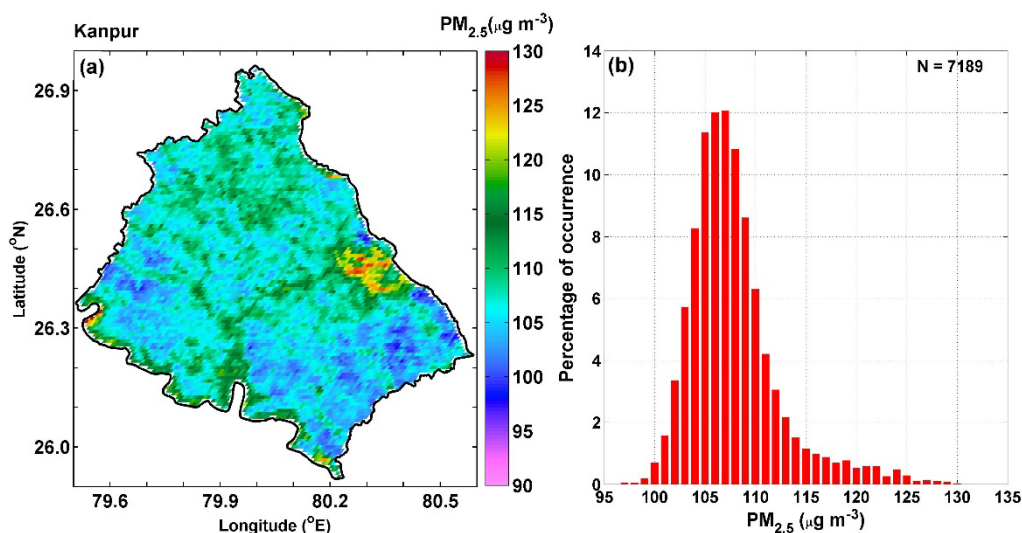


Figure 9: The spatial map of the annual mean $PM_{2.5}$ over the Kanpur region and its frequency distribution

The seasonal mean of $PM_{2.5}$ concentrations over Kanpur is shown in Figure 10. Among seasons, WIN recorded the highest $PM_{2.5}$ ($141 \pm 9.4 \mu\text{g m}^{-3}$) followed by PMN ($138 \pm 6.7 \mu\text{g m}^{-3}$), SUM ($81 \pm 3.3 \mu\text{g m}^{-3}$), and MON ($71 \pm 5.6 \mu\text{g m}^{-3}$). The seasonal variation of $PM_{2.5}$ over Kanpur is similar to that observed over Delhi-NCR. The seasonal mean $PM_{2.5}$ values were slightly higher over Kanpur compared to Delhi-NCR.

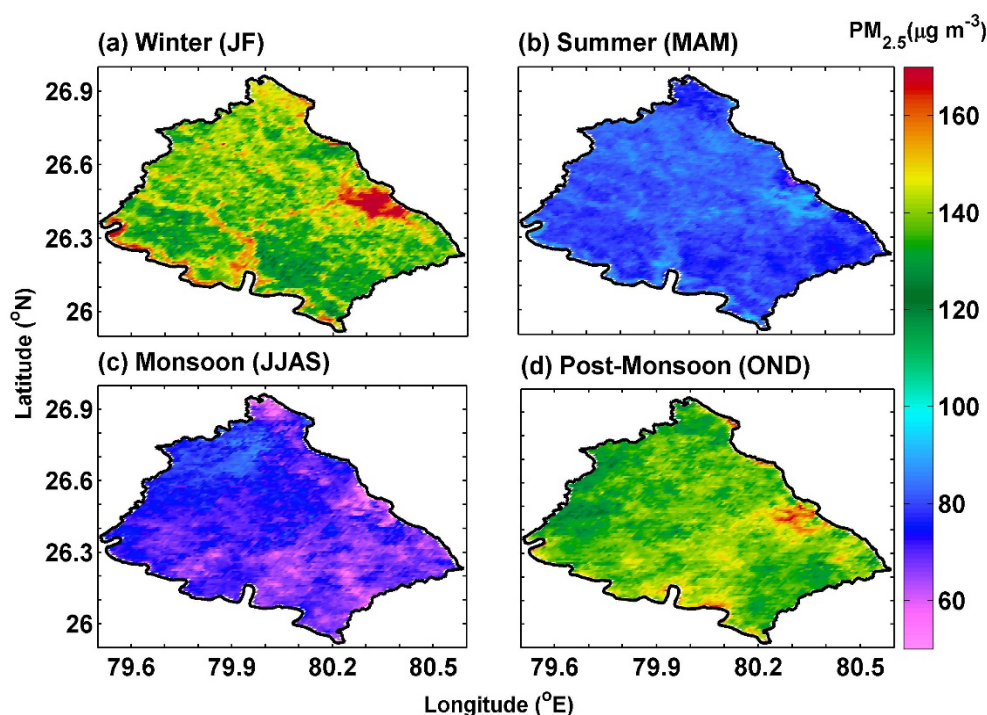


Figure 10: The seasonal mean $PM_{2.5}$ maps over the Kanpur region

7.3.3 Bengaluru region

The gridded annual mean $PM_{2.5}$ over the Bengaluru study region is shown in Figure 11(a). The Bengaluru region includes urban, peri-urban, and rural areas (Bengaluru urban and Bengaluru rural districts). Compared to Kanpur and Delhi-NCR, $PM_{2.5}$ over Bengaluru was lower by almost three times. The annual mean concentrations varied between 35 and $55 \mu\text{g m}^{-3}$ with maximum concentrations ($\sim 55 \mu\text{g m}^{-3}$) recorded over Bengaluru urban. The Bengaluru annual mean $PM_{2.5}$

values again could be an overestimation because of the missing MON $PM_{2.5}$ estimation due to corresponding AOD data gaps. The number of MON days with missing AOD data over Bengaluru was much higher than that of Delhi-NCR and the Kanpur region.

Based on the CAAQMS year-round data and model-predicted corresponding gridded daily mean $PM_{2.5}$ values, the Bengaluru model-predicted annual mean $PM_{2.5}$ was found to be ~30% to 35% overestimated. In the case of Bengaluru, the emphasis was more on seasonal maps rather than the annual map. The spatially averaged annual mean $PM_{2.5}$ for the study year was $44 \pm 2.5 \mu\text{g m}^{-3}$. Figure 11(b) shows the frequency distribution of the annual mean $PM_{2.5}$ concentrations. In total, 5,244 grids (1 km size) cover the Bengaluru urban, peri-urban, and rural areas. The frequency distribution exhibits a modal $PM_{2.5}$ value of $\sim 43 \mu\text{g m}^{-3}$. Over the Bengaluru region, ~98% of the gridded annual mean $PM_{2.5}$ values exceeded the national annual standard ($40 \mu\text{g m}^{-3}$).

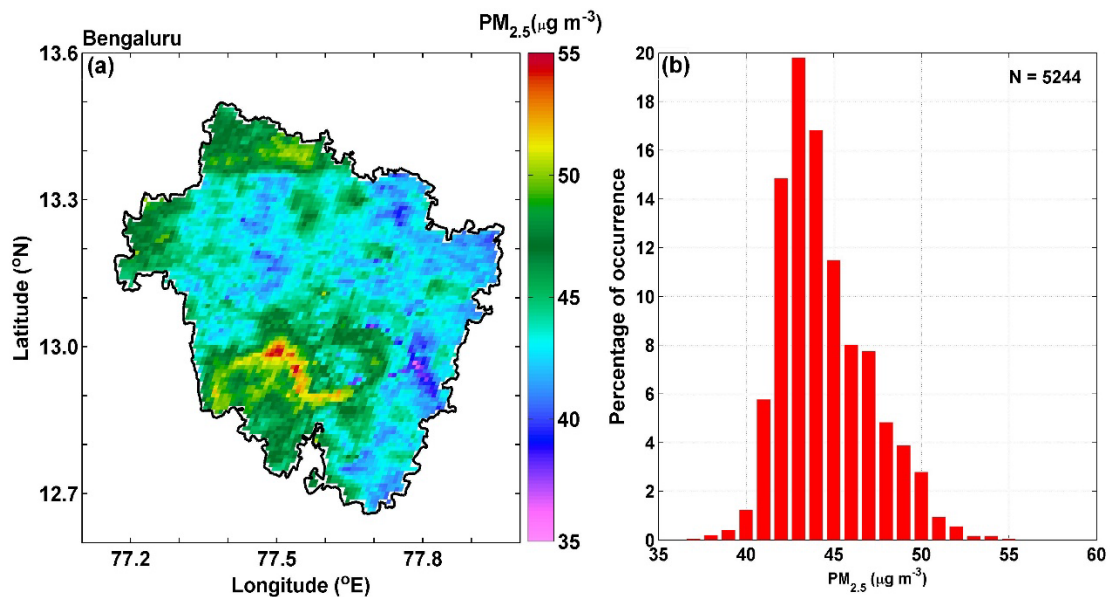


Figure 11: The spatial map of the annual mean $PM_{2.5}$ over the Bengaluru region and its frequency distribution

Figure 12 depicts the seasonal mean maps of $PM_{2.5}$ over the Bengaluru region. As expected, the MON concentrations were the lowest, WIN saw the highest, and PMN $PM_{2.5}$ was moderate. Spatial gaps in $PM_{2.5}$ during MON because of the missing AOD is evident from the figure. The highest mean $PM_{2.5}$ values were noted in WIN ($49 \pm 2.8 \mu\text{g m}^{-3}$) followed by SUM ($44 \pm 2.5 \mu\text{g m}^{-3}$), PMN ($34 \pm 3.6 \mu\text{g m}^{-3}$), and MON ($16 \pm 4.0 \mu\text{g m}^{-3}$) seasons (Table 4). Over Bengaluru, SUM $PM_{2.5}$ values followed WIN values, while over Delhi-NCR and Kanpur, PMN values followed WIN values. This pattern was observed in CAAQMS $PM_{2.5}$ measurements also (Figure 3).

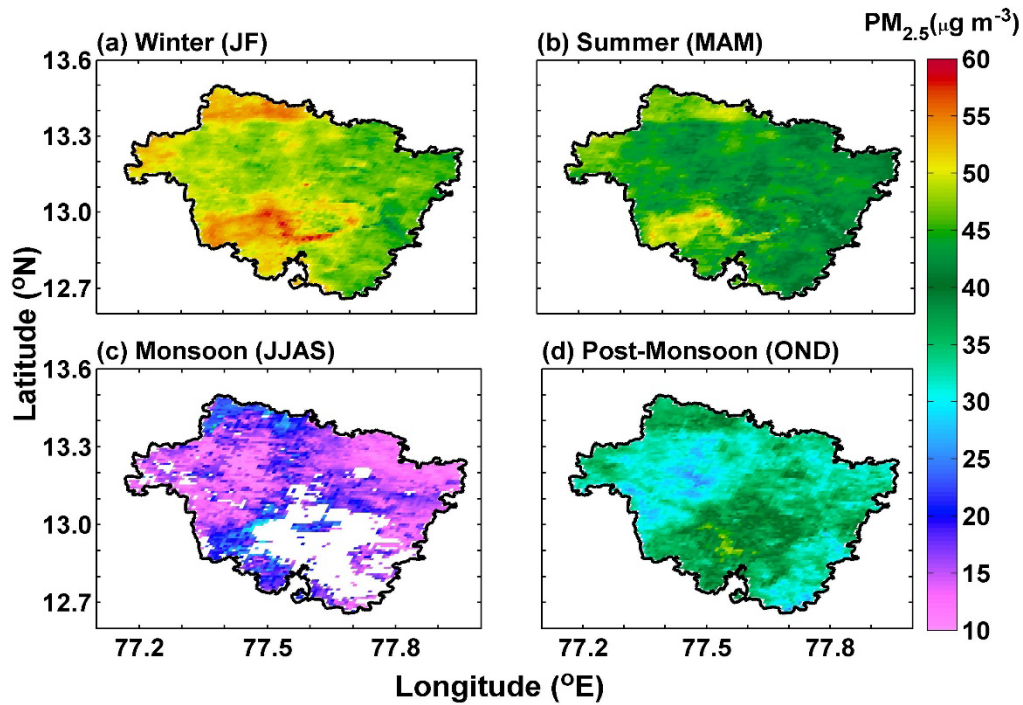


Figure 12: The seasonal mean PM_{2.5} maps over the Bengaluru region

Table 4: Annual and seasonal mean PM_{2.5} statistics. IQR and SD represent the interquartile range and standard deviation, respectively.

Season	Delhi-NCR		Kanpur region		Bengaluru region	
	Mean (SD) ($\mu\text{g m}^{-3}$)	Median (IQR) ($\mu\text{g m}^{-3}$)	Mean (SD) ($\mu\text{g m}^{-3}$)	Median (IQR) ($\mu\text{g m}^{-3}$)	Mean (SD) ($\mu\text{g m}^{-3}$)	Median (IQR) ($\mu\text{g m}^{-3}$)
WIN	134 (16.0)	133 (25.8)	141 (9.4)	139 (11.3)	49 (2.8)	48 (4.0)
SUM	79 (5.3)	79 (7.0)	81 (3.3)	81 (4.4)	44 (2.5)	43 (3.0)
MON	57 (8.2)	58 (12.0)	71 (5.6)	70 (7.0)	16 (4.0)	15 (5.0)
PMN	131(11.9)	130 (15.7)	138 (6.7)	137 (7.2)	34 (3.6)	34 (5.0)
Annual	101 (7.5)	100 (10.5)	108 (4.5)	107 (4.5)	44 (2.5)	44 (3.3)

7.3.4 Urban areas of Delhi, Kanpur, and Bengaluru

Spatial maps of the annual and seasonal mean PM_{2.5} for the urban areas of Delhi, Kanpur, and Bengaluru are shown in this subsection.

The municipal corporation of Delhi-NCT is divided into 11 revenue districts (North, North East, North West, West, South, South West, South East, New Delhi, Central, Shahdara, and East) and 290 wards. Figures 13 and 14 show the annual and seasonal mean spatial distribution of PM_{2.5} over Delhi-NCT, overlaid with revenue district and ward boundaries. The estimated annual mean \pm SD (median \pm IQR) concentration of PM_{2.5} across Delhi-NCT for the study year was $117 \pm 6.8 \mu\text{g m}^{-3}$ ($116 \pm 11.0 \mu\text{g m}^{-3}$). The gridded maximum and minimum mean PM_{2.5} values were

~102 and 134 $\mu\text{g m}^{-3}$, respectively. The seasonal and annual mean $\text{PM}_{2.5}$ for the 11 revenue districts are given in Table 5. The highest annual mean $\text{PM}_{2.5}$ estimated was for the Shahdara district (~126 $\mu\text{g m}^{-3}$) followed by the East (~124 $\mu\text{g m}^{-3}$), West (~122 $\mu\text{g m}^{-3}$), South East (~120 $\mu\text{g m}^{-3}$), North East and Central (~119 $\mu\text{g m}^{-3}$), North West (~118 $\mu\text{g m}^{-3}$), New Delhi (~116 $\mu\text{g m}^{-3}$), South (~115 $\mu\text{g m}^{-3}$), North (~114 $\mu\text{g m}^{-3}$), and South West (~113 $\mu\text{g m}^{-3}$) districts.

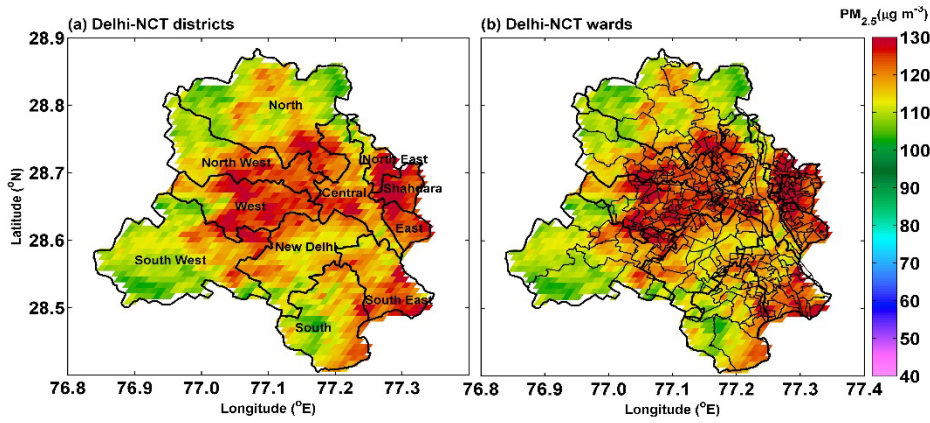


Figure 13: The spatial map of the annual mean $\text{PM}_{2.5}$ over Delhi-NCT region with overlaid revenue district and ward boundaries

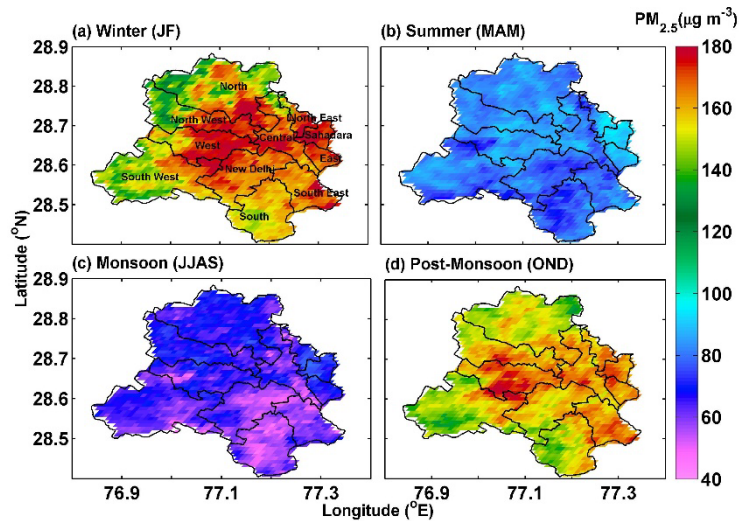


Figure 14: Spatial maps of the seasonal mean $\text{PM}_{2.5}$ over Delhi-NCT

Table 5: Annual and seasonal statistics of $\text{PM}_{2.5}$ for various Delhi-NCT revenue districts. IQR and SD represent the interquartile range and standard deviation, respectively.

Delhi-NCT revenue districts	Season	Minimum ($\mu\text{g m}^{-3}$)	Maximum ($\mu\text{g m}^{-3}$)	Mean (SD) ($\mu\text{g m}^{-3}$)	Median (IQR) ($\mu\text{g m}^{-3}$)
East	WIN	158	179	169 (4.4)	170 (5.5)
	SUM	76	95	88 (4.0)	88 (5.1)
	MON	47	79	67 (5.9)	67 (6.7)
	PMN	154	174	164 (4.7)	164 (6.3)
	Annual	113	133	124 (3.9)	123 (4.5)

North East	WIN	126	183	166 (13.2)	170 (13.8)
	SUM	67	96	85 (8.1)	85 (14.6)
	MON	44	80	67 (8.1)	66 (13.4)
	PMN	142	173	160 (6.7)	161 (9.8)
	Annual	103	132	119 (8.2)	121 (14.7)
Central	WIN	126	183	166 (10.1)	169 (10.0)
	SUM	67	92	82 (5.0)	82 (7.0)
	MON	40	77	62 (6.6)	62 (9.6)
	PMN	142	177	161 (6.2)	161 (8.7)
	Annual	103	131	119 (5.7)	119 (7.6)
New Delhi	WIN	149	181	166 (6.1)	166 (8.0)
	SUM	69	88	78 (4.0)	78 (5.7)
	MON	48	72	58 (5.4)	58 (8.0)
	PMN	137	174	156 (7.3)	156 (9.6)
	Annual	106	128	116 (4.6)	116 (6.3)
North	WIN	123	187	155 (14.9)	156 (25.0)
	SUM	76	94	83 (3.2)	83 (4.1)
	MON	47	78	67 (4.3)	67 (5.9)
	PMN	130	171	152 (7.3)	151 (9.8)
	Annual	102	130	114 (6.3)	113 (9.5)
North West	WIN	121	182	158 (16.7)	162 (31.2)
	SUM	77	93	84 (3.1)	84 (3.8)
	MON	51	77	68 (4.3)	68 (6.5)
	PMN	142	188	159 (7.7)	158 (10.9)
	Annual	105	132	118 (6.4)	118 (11.4)
Shahdara	WIN	166	183	174 (4.2)	174 (6.2)
	SUM	78	96	91 (3.5)	91 (3.4)
	MON	59	80	73 (4.6)	73 (6.1)
	PMN	154	174	165 (5.1)	164 (7.9)
	Annual	117	134	126 (3.8)	126 (5.5)
South	WIN	142	172	157 (5.4)	157 (5.7)
	SUM	67	88	78 (4.5)	78 (6.0)

	MON	44	70	57 (5.1)	57 (7.2)
	PMN	136	173	152 (8.0)	151 (12.9)
	Annual	102	129	115 (5.4)	115 (7.7)
South East	WIN	154	195	167 (7.3)	165 (8.0)
	SUM	70	90	80 (4.4)	80 (5.9)
	MON	40	69	58 (5.4)	58 (6.8)
	PMON	150	176	163 (5.7)	163 (7.9)
	Annual	109	133	120 (5.2)	120 (6.8)
South West	WIN	132	181	154 (11.5)	152 (16.4)
	SUM	62	92	80 (5.5)	80 (7.6)
	MON	46	74	62 (4.7)	63 (5.7)
	PMON	132	188	151 (10.1)	148 (11.6)
	Annual	102	131	113 (6.2)	111 (7.6)
West	WIN	141	181	169 (10)	173 (12.2)
	SUM	66	92	83 (5.3)	84 (5.3)
	MON	43	76	65 (5.4)	65 (6.2)
	PMON	138	188	166 (8.9)	167 (10.5)
	Annual	105	131	122 (6.1)	124 (7.4)

Kanpur urban covers an approximate area of 260 square kilometres. The Kanpur Municipal Corporation (KMC) is divided into six zones and 110 wards. All the zones except Zone 6 have 18 wards each. Annual and seasonal mean spatial maps of PM_{2.5} for the study year are shown in Figures 15 and 16 (with zone and ward boundaries overlaid). The maximum (minimum) annual mean value of PM_{2.5} was ~128 (97) $\mu\text{g m}^{-3}$. The spatially averaged annual mean PM_{2.5} value for Kanpur urban was $115 \pm 7.1 \mu\text{g m}^{-3}$. The zone-wise annual and seasonal PM_{2.5} statistics are given in Table 6. Zone 1 annual PM_{2.5} had the highest values ($122 \pm 3.9 \mu\text{g m}^{-3}$), while Zone 2 had the lowest ($112 \pm 7.5 \mu\text{g m}^{-3}$). It has to be noted that the Panki Thermal Power Station is located in Zone 5, which is a coal-based thermal power plant having a total generating capacity of 210 MW.

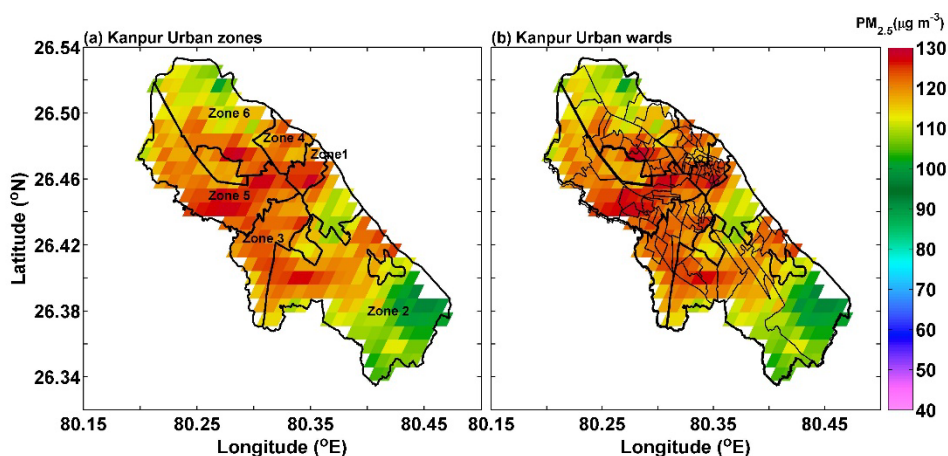


Figure 15: The spatial map of the annual mean PM_{2.5} over Kanpur urban (zone and ward boundaries overlaid)

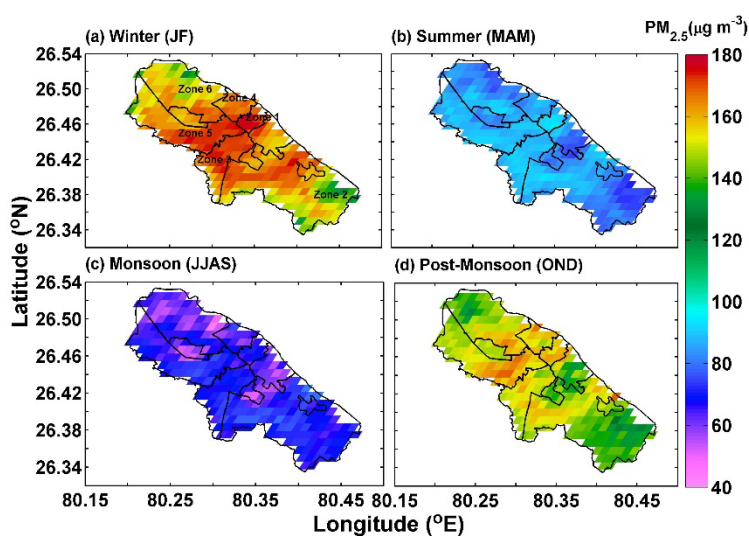


Figure 16: Spatial maps of the seasonal mean PM_{2.5} over Kanpur urban

Table 6: Annual and seasonal statistics of PM_{2.5} for various Kanpur urban zones. IQR and SD represent the interquartile range and standard deviation, respectively.

Kanpur urban zones	Season	Minimum (µg m ⁻³)	Maximum (µg m ⁻³)	Mean (SD) (µg m ⁻³)	Median (IQR) (µg m ⁻³)
Zone 1	WIN	163	177	172 (4.3)	172 (6.9)
	SUM	80	95	89 (3.9)	90 (5.7)
	MON	58	71	66 (3.5)	66 (4.5)
	PMN	148	166	157 (5.1)	158 (6.7)
	Annual	114	128	122 (3.9)	123 (5.8)
Zone 2	WIN	123	175	156 (11.4)	156 (17.2)
	SUM	71	94	83 (5.4)	82 (8.7)
	MON	50	84	70 (5.6)	70 (6.6)
	PMN	126	168	144 (8.0)	143 (11.6)

	Annual	97	127	112 (7.5)	111 (11.5)
Zone 3	WIN	145	174	169 (8.0)	169 (11.1)
	SUM	79	92	88 (3.0)	89 (4.5)
	MON	52	76	68 (4.1)	68 (5.4)
	PMN	139	161	150 (5.4)	150 (8.3)
	Annual	109	125	119 (4.1)	119 (6.8)
Zone 4	WIN	160	177	167 (5.0)	167 (7.0)
	SUM	80	93	87 (3.6)	86 (5.1)
	MON	56	76	65 (5.1)	65 (6.7)
	PMN	147	166	157 (3.9)	157 (4.4)
	Annual	114	125	121 (3.4)	121 (4.6)
Zone 5	WIN	135	175	165 (9.9)	165 (14.8)
	SUM	81	92	88 (2.7)	88 (3.5)
	MON	52	77	68 (4.6)	68 (5.9)
	PMN	140	169	153 (6.9)	152 (11.0)
	Annual	108	128	120 (4.9)	121 (7.0)
Zone 6	WIN	129	173	159 (9.0)	159 (11.4)
	SUM	73	93	84 (4.8)	84 (5.9)
	MON	43	76	62 (5.7)	62 (7.4)
	PMN	129	168	149 (8.4)	149 (12.4)
	Annual	101	128	116 (6.2)	116 (10.1)

The Bruhat Bengaluru Mahanagara Palike (BBMP) has divided Bengaluru urban into 198 wards and eight zones. The zones are Yelahanka, Dasarahalli, Rajarajeshwari Nagar, Bommanahalli, South, West, East, and Mahadevapura. The spatial distribution of the annual and seasonal mean PM_{2.5} for the study year are shown in Figures 17 and 18. Annual and seasonal PM_{2.5} statistics for the Bengaluru urban zones are summarised in Table 7. Dasarahalli, Rajarajeshwari Nagar, Bommanahalli, South, and West zones exhibited considerable higher PM_{2.5} values (Figure 17). Rajarajeshwari Nagar ($\sim 49 \pm 2.2 \mu\text{g m}^{-3}$) and South ($49 \pm 2.9 \mu\text{g m}^{-3}$) zones had the maximum annual PM_{2.5} values followed by the West ($48 \pm 2.6 \mu\text{g m}^{-3}$), Bommanahalli ($48 \pm 2.7 \mu\text{g m}^{-3}$), and Dasarahalli ($48 \pm 2.9 \mu\text{g m}^{-3}$) zones. The Mahadevapura zone had the lowest with an annual mean PM_{2.5} of $44 \pm 2.5 \mu\text{g m}^{-3}$ for the study year. In the case of Bengaluru, the seasonal mean PM_{2.5} is more representative compared to that of annual mean values, as discussed in earlier sections.

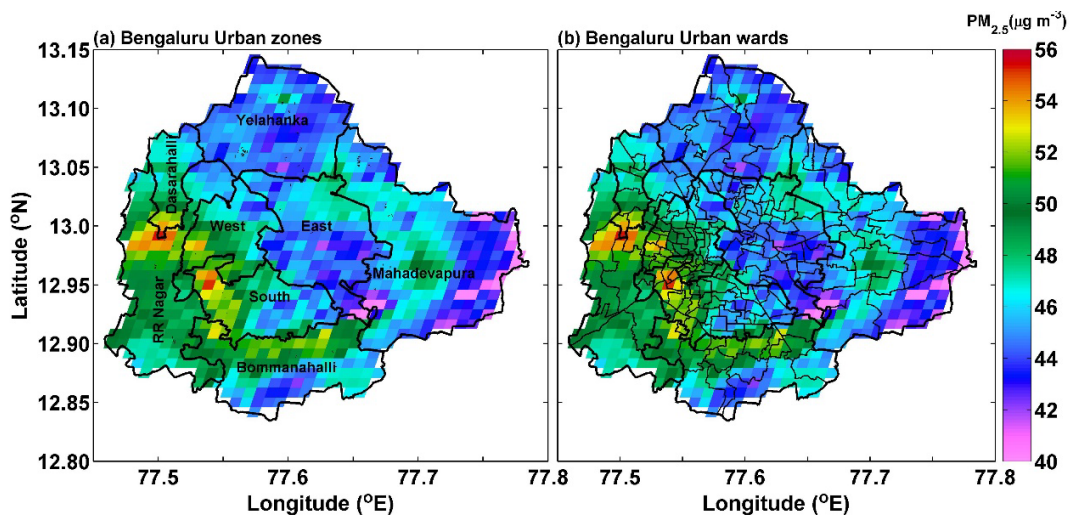


Figure 17: The spatial map of the annual mean PM_{2.5} over Bengaluru urban (zone and ward boundaries overlaid)

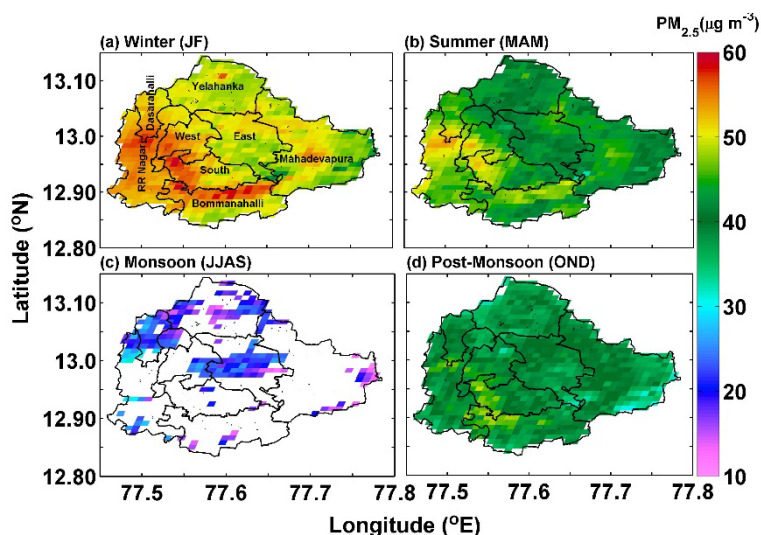


Figure 18: Spatial maps of the seasonal mean PM_{2.5} over Bengaluru urban

Table 7: Annual and seasonal statistics of PM_{2.5} for various Bengaluru urban zones. IQR and SD represent the interquartile range and standard deviation, respectively.

Bengaluru urban zones	Season	Minimum ($\mu\text{g m}^{-3}$)	Maximum ($\mu\text{g m}^{-3}$)	Mean (SD) ($\mu\text{g m}^{-3}$)	Median(IQR) ($\mu\text{g m}^{-3}$)
Bommanahalli	WIN	43	59	52(3.2)	51 (5.0)
	SUM	36	50	45 (2.4)	45 (3.4)
	MON	13	21	17 (2.5)	17 (2.6)
	PMN	33	48	41 (3.8)	40 (5.9)
	Annual	41	53	48 (2.7)	47 (4.5)
Yelahanka	WIN	46	56	49(1.4)	49 (1.8)
	SUM	36	47	42 (1.8)	42 (2.3)
	MON	13	25	19 (3.1)	19 (5.4)

	PMN	32	43	38 (2.1)	38 (2.8)
	Annual	42	51	45 (1.3)	45 (1.3)
Dasarahalli	WIN	49	58	53 (2.4)	53 (3.7)
	SUM	41	55	47 (3.1)	46 (3.7)
	MON	15	31	22 (3.8)	23 (4.1)
	PMN	34	47	40 (3.1)	40 (4.9)
	Annual	44	55	48 (2.9)	47 (3.5)
East	WIN	42	54	50 (1.9)	50 (2.6)
	SUM	33	50	42 (2.3)	42 (3.2)
	MON	12	26	21 (3.5)	22 (4.7)
	PMN	32	44	39 (2.0)	39 (2.6)
	Annual	38	49	45 (1.8)	45 (2.4)
Mahadevapura	WIN	40	55	49 (2.7)	49 (3.3)
	SUM	32	47	42 (2.4)	42 (2.6)
	MON	5	25	16 (4.5)	15 (5.3)
	PMN	28	44	38 (3.0)	38 (4.4)
	Annual	36	50	44 (2.5)	45 (3.9)
Rajarajeshwari Nagar	WIN	49	58	53 (1.8)	53 (2.6)
	SUM	39	55	47 (2.8)	47 (3.6)
	MON	13	32	23 (4.5)	23 (7.2)
	PMN	34	42	40 (3.2)	40 (4.9)
	Annual	44	55	49 (2.2)	49 (2.9)
South	WIN	47	59	53 (2.9)	53 (4.8)
	SUM	41	52	46 (3.0)	46 (5.0)
	MON	14	21	19 (2.7)	19 (4.0)
	PMN	36	48	42 (3.3)	42 (4.9)
	Annual	43	55	49 (2.9)	49 (5.0)
West	WIN	48	59	53 (2.3)	53 (3.5)
	SUM	36	53	45 (3.9)	46 (5.7)
	MON	17	27	23 (2.4)	23 (2.9)
	PMN	33	47	41 (2.8)	41 (3.7)
	Annual	43	54	48 (2.6)	48 (3.7)

7.4 PM_{2.5} hotspots

7.4.1 Delhi-NCR region

PM_{2.5} hotspot areas over Delhi-NCR are shown in Figure 19 (left panel). The hotspot analysis shown in the figure was carried out on the annual mean map of PM_{2.5}. Note that the hotspot identification is subject to the spatial extent of the study area. In Delhi-NCR, most of the NCT region and its surrounding areas were identified as hotspots (with the highest confidence level). Scattered hotspots were also identified in the districts of Nuh, Bharatpur, Meerut, and Sonipat. The right panel of Figure 19 shows the GHSL-SMOD settlement classification. Most of the hotspot areas coincided with urban areas. Within Delhi-NCT, all the revenue districts were identified as hotspot areas.

PM_{2.5} hotspots were also identified based on seasonal mean maps of PM_{2.5} (Figure 20). During WIN and PMN, hotspots were spread over Delhi-NCT and surrounding districts. No pattern was observed in MON hotspots. The PM_{2.5} levels in the hotspot areas can be different across seasons. During SUM, hotspot areas were also observed in the Bharatpur district of Rajasthan. This could be because of the frequent summertime dust storms over that region. Within Delhi-NCT, during WIN, the entire districts of North East, Central, and Shahdara and parts of East, New Delhi, North, North West, South, South East, South West, and West were identified as hotspots. During SUM, the entire district of Shahdara and parts of East, North East, Central, North, North West, South West, and West were identified as hotspots. During PMN, the entire districts of East, North East, Central, New Delhi, Shahdara, South-East, West and parts of New Delhi, North, North West, South, and South West were identified as hotspots.

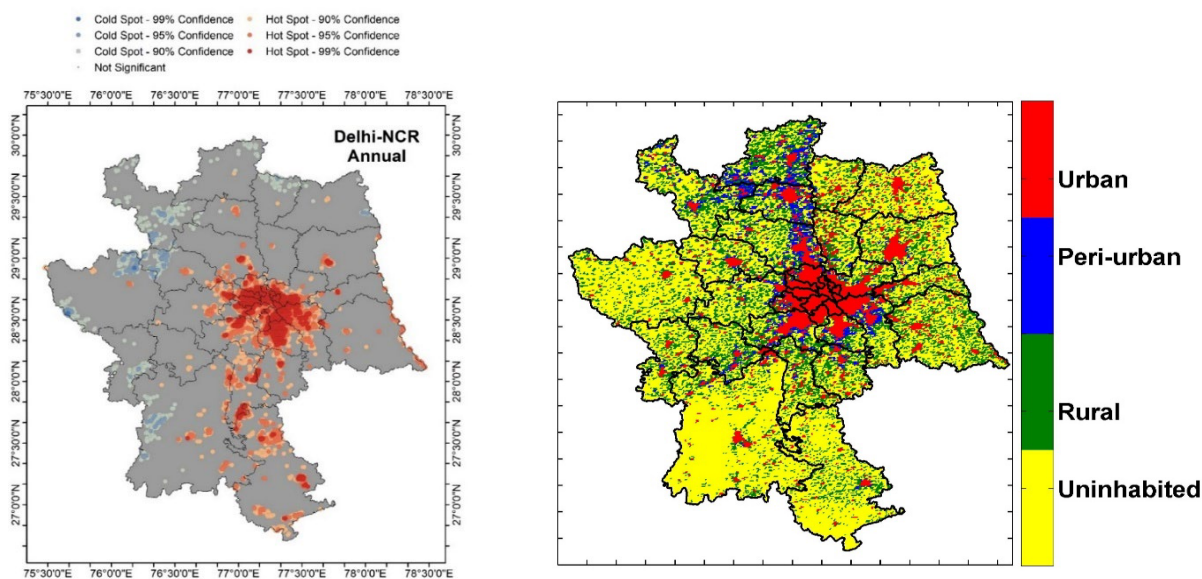


Figure 19: Annual hotspot analysis for Delhi-NCR. The right panel corresponds to the GHSL-SMOD settlement classification.

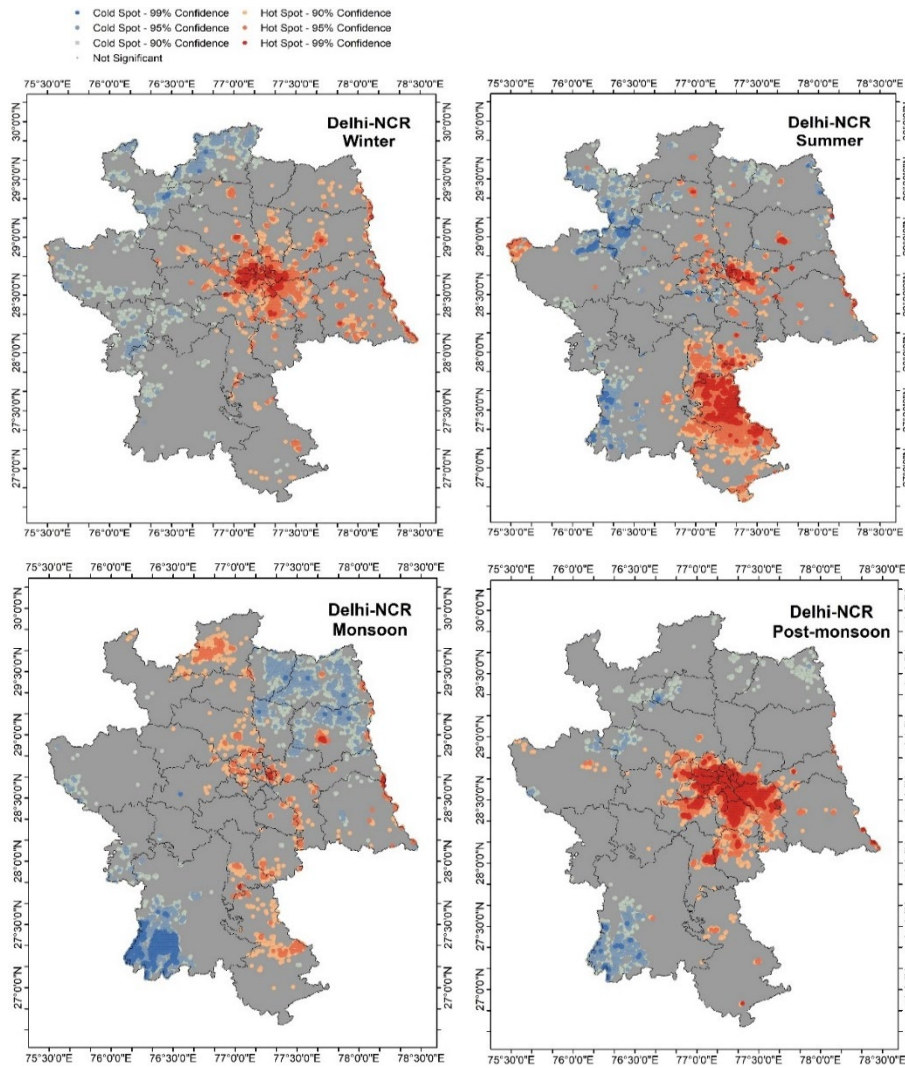


Figure 20: Seasonal hotspot analysis for Delhi-NCR

7.4.2 Kanpur region

Figure 21 (left panel) presents the hotspot analysis for the Kanpur region. Hotspot areas were observed in the Kanpur urban district, particularly in Kanpur city. All the zones (except parts of Zones 2 and 6) of Kanpur city were identified as hotspot areas. PM_{2.5} hotspots were observed along the river basins also.

WIN hotspot areas resembled the annual hotspot areas. During SUM, the entire Zone 1 and parts of all other zones were identified as hotspot areas. During PMN, the entire Zones 1 and 4 and parts of Zones 2, 3, and 5 were observed as hotspot areas. For MON, no grids were identified as hotspots with the highest confidence level.

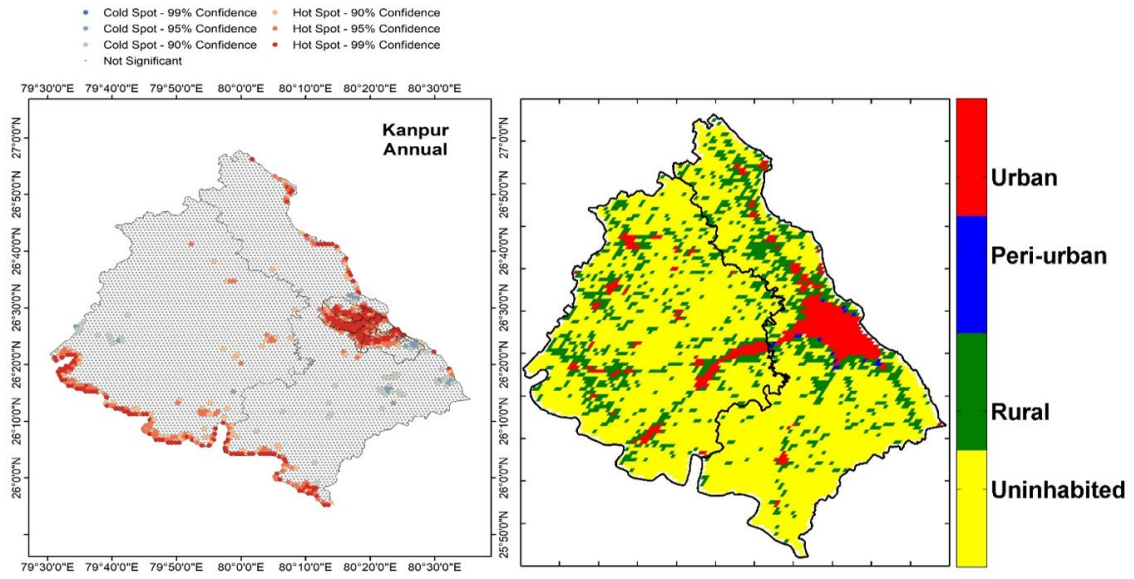


Figure 21: Annual hotspot analysis for the Kanpur region

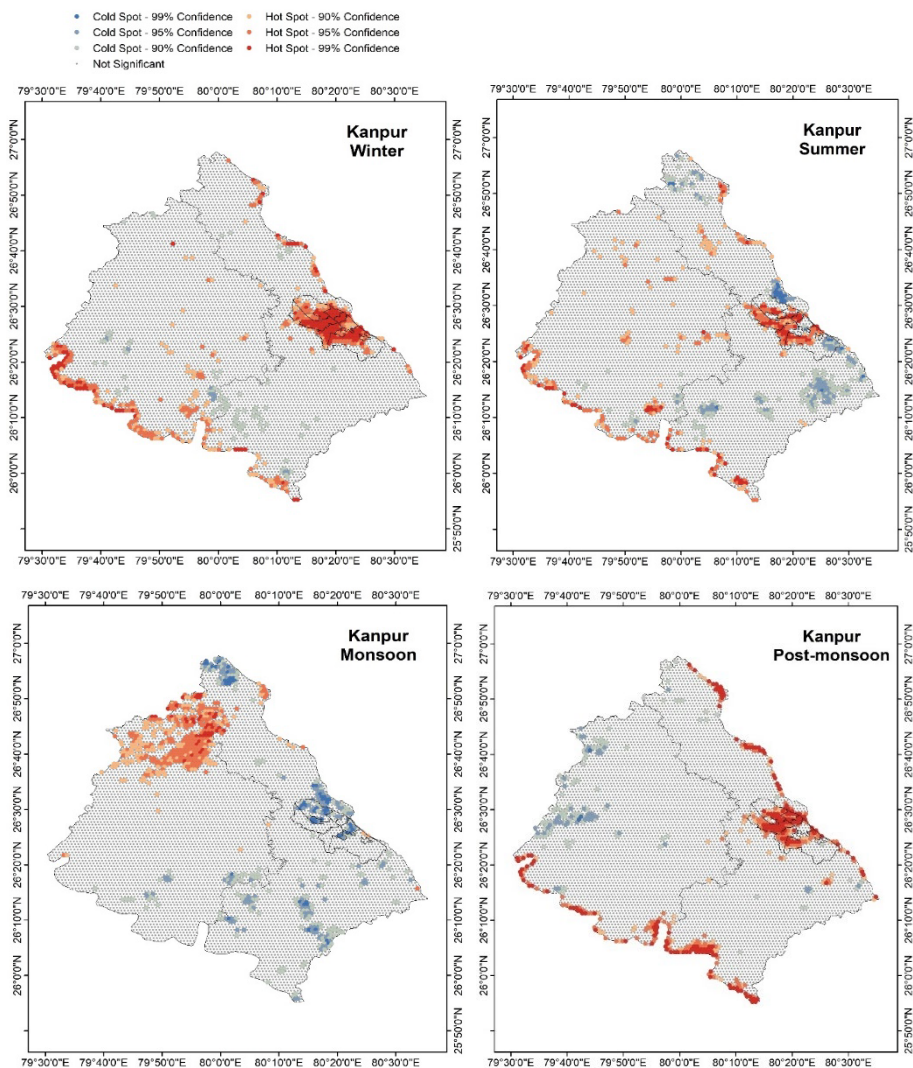


Figure 22: Seasonal hotspot analysis for the Kanpur region

7.4.3 Bengaluru region

Annual and seasonal hotspot analyses for the Bengaluru region are shown in Figures 23 and 24. Most of the hotspot areas were within the urban district of Bengaluru. Parts of Bommanahalli, Dasarahalli, Rajarajeshwari Nagar, and South and West zones were observed as hotspot areas on annual and seasonal scales (except MON). During SUM, hotspot areas were observed in the Greater Bengaluru region also.

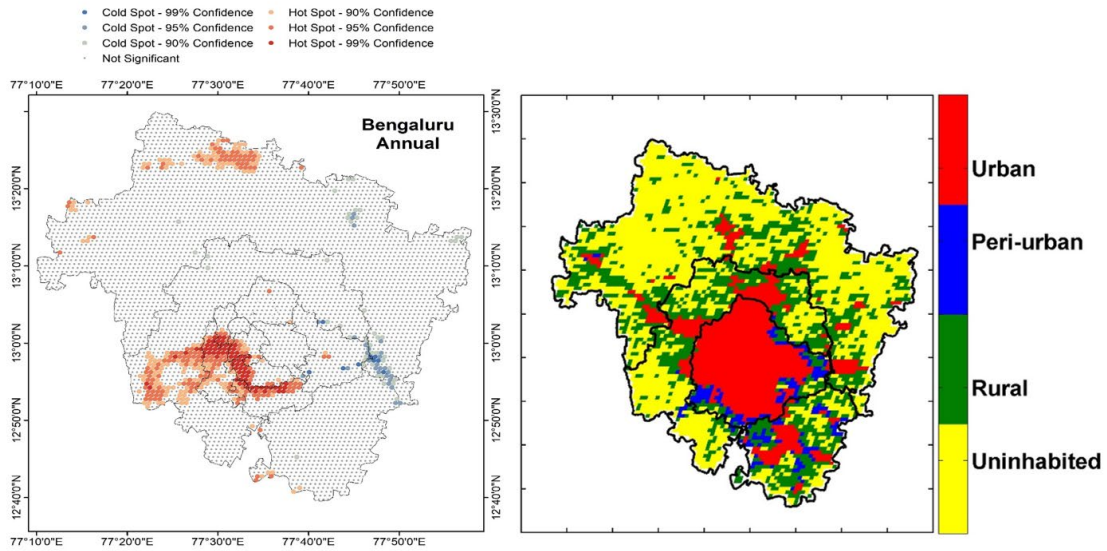


Figure 23: Annual hotspot analysis for the Bengaluru region

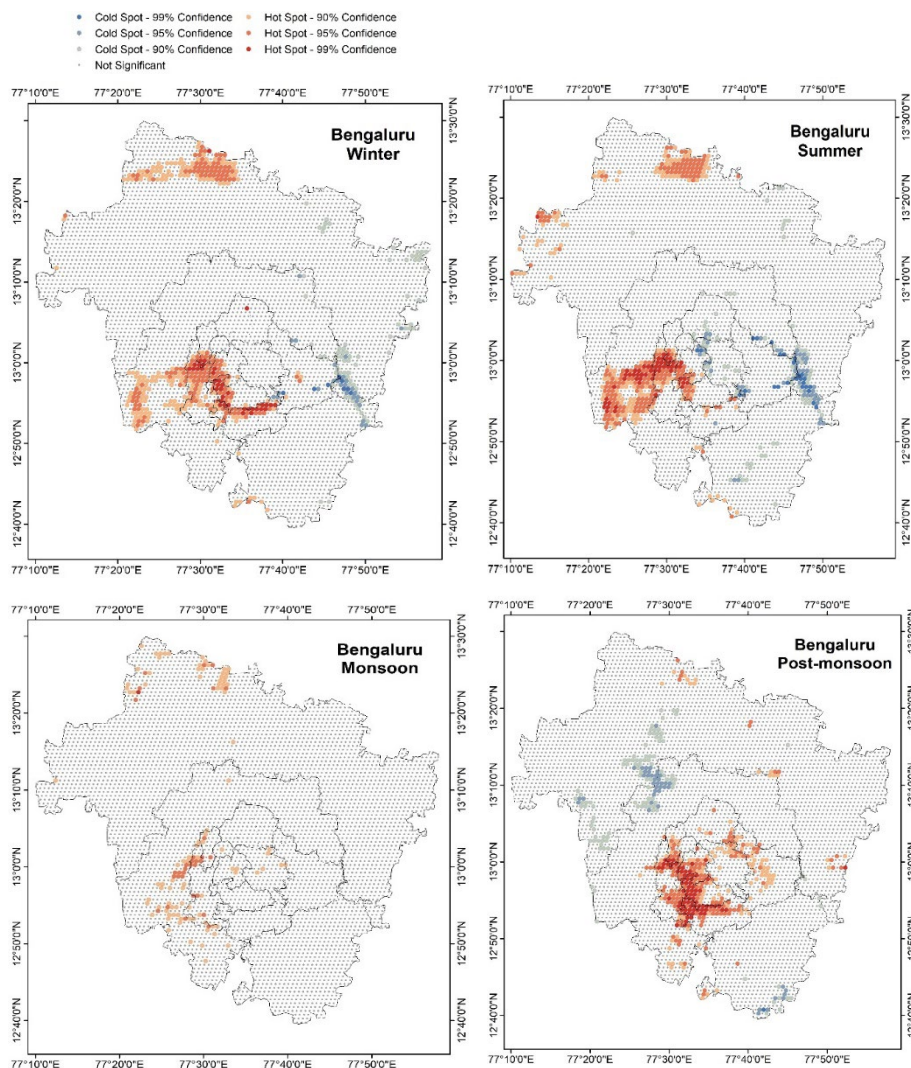


Figure 24: Seasonal hotspot analysis for the Bengaluru region

7.5 Spatial gradients

To understand the rural, peri-urban, and urban contrast in $PM_{2.5}$ (over the study regions), $PM_{2.5}$ estimates over the GHSL-SMOD identified different settlement classes were aggregated and plotted at annual and seasonal scales.

7.5.1 Delhi-NCR region

For this region, the uninhabited, rural, peri-urban, and urban areas constituted $\sim 59.8\%$, 24.0% , 4.6% , and 11.6% , respectively. The box plot (see Figure 25) depicts the variation in the annual mean $PM_{2.5}$ for different settlement classes/areas of Delhi-NCR. Spatially, the highest mean values were observed for urban settlements ($109 \pm 9.4 \mu g m^{-3}$) and the lowest ($100 \pm 6.6 \mu g m^{-3}$) for the uninhabited class. The rural and uninhabited $PM_{2.5}$ mean, median, maximum, minimum, and percentiles values were comparable. The spatial gradients in the annual mean $PM_{2.5}$ were not as steep as expected. Peri-urban and urban settlement classes exhibited a mere difference of $\sim 7 \mu g m^{-3}$.

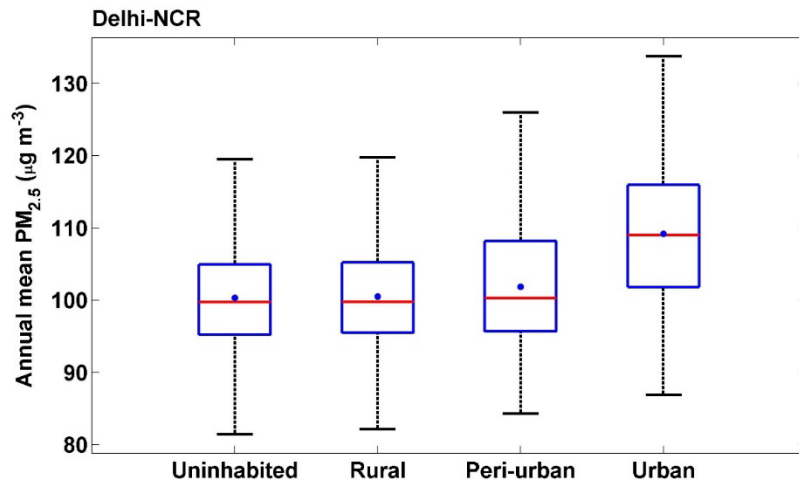


Figure 25: Variations in the annual mean PM_{2.5} for Delhi-NCR for the identified settlement classes: uninhabited, rural, peri-urban, and urban

Figure 26 shows variation in the seasonal mean PM_{2.5} for different settlement classes of Delhi-NCR. Seasonally, the WIN and PMN mean PM_{2.5} values were higher, while MON observed the lowest PM_{2.5} for all settlement classes. The annual and seasonal mean PM_{2.5} statistics for the different settlement classes of Delhi-NCR are given in Table 8. SUM and MON variations were shallower compared to other seasons.

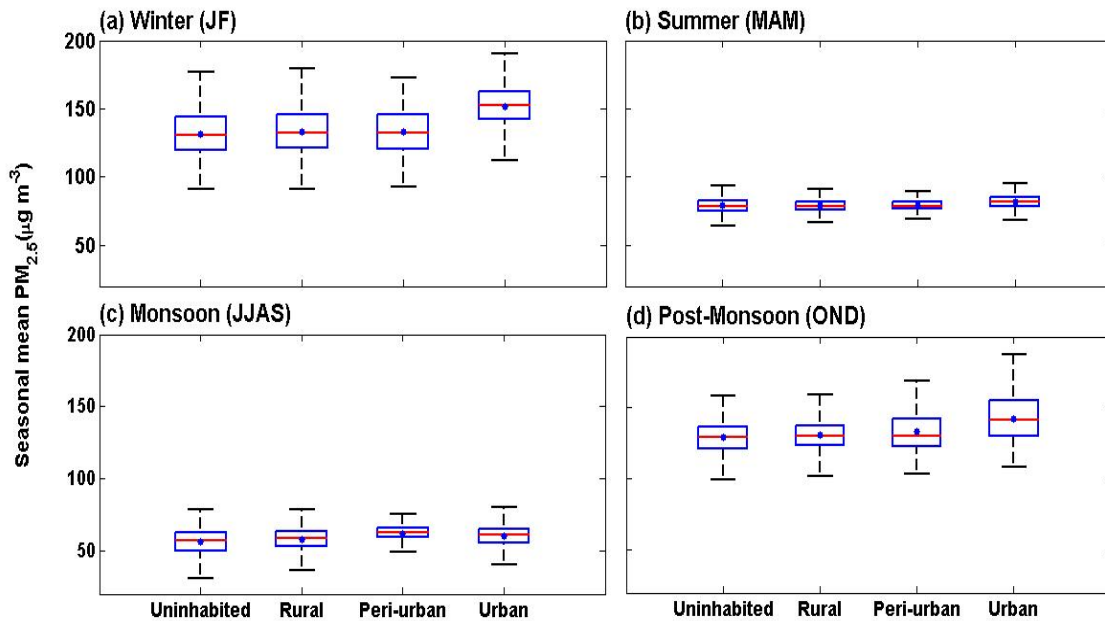


Figure 26: Variations in the seasonal mean PM_{2.5} for Delhi-NCR for the identified settlement classes

7.5.2 Kanpur region

Figure 27 shows the variations in the annual mean PM_{2.5} between various settlement classes of the Kanpur region. For the Kanpur region, the uninhabited, rural, peri-urban, and urban areas constituted ~73.0%, 19.8%, 0.2%, and 6.9%, respectively. From the figure, it can be observed that the urban area was characterised by the highest mean PM_{2.5} ($114 \pm 6.4 \mu\text{g m}^{-3}$), followed by peri-urban ($112 \pm 4.4 \mu\text{g m}^{-3}$), rural ($108 \pm 4.6 \mu\text{g m}^{-3}$), and uninhabited ($107 \pm 3.7 \mu\text{g m}^{-3}$) settlement classes.

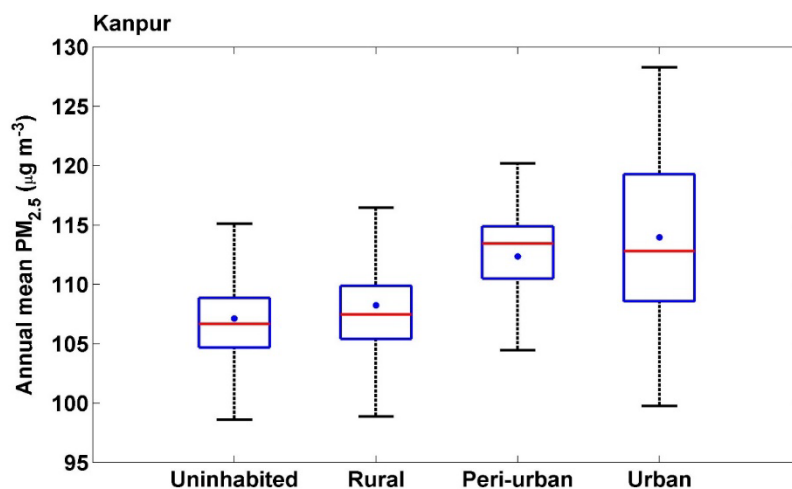


Figure 27: Variations in the annual mean PM_{2.5} for the Kanpur region for the identified settlement classes

The seasonal mean of PM_{2.5} values over Kanpur for different settlement classes are shown in Figure 28. Similar to that observed over Delhi-NCR, SUM and MON gradients were shallower. MON PM_{2.5} remained the lowest with no variations across settlement classes. Statistics are given in Table 8.

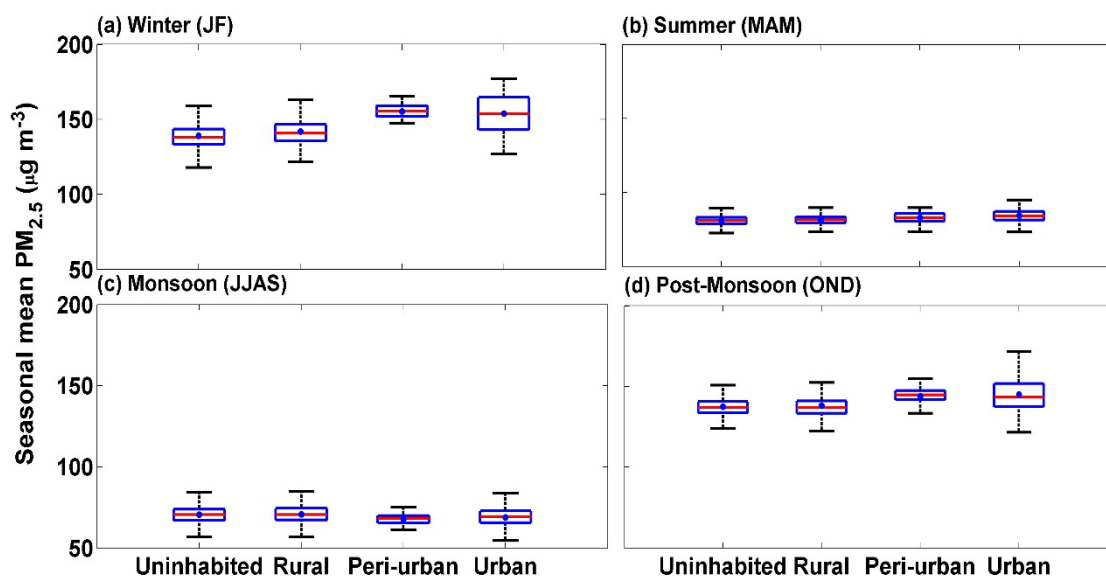


Figure 28: Variations in the seasonal mean PM_{2.5} of the Kanpur region for the identified settlement classes

7.5.3 Bengaluru region

The PM_{2.5} values over the Bengaluru region were almost three times lower compared to Delhi-NCR and Kanpur region. For the Bengaluru region, the uninhabited, rural, peri-urban, and urban areas constituted ~49.6%, 27%, 3.7%, and 19.7%, respectively. Figure 29 depicts the variation in annual PM_{2.5} values for different settlement classes across the Bengaluru region. The annual observed mean PM_{2.5} values for urban, peri-urban, rural, and uninhabited settlements were ~46, 45, 44, and 44 µg m⁻³, respectively. The rural-urban contrast in annual PM_{2.5} was hardly observed. Seasonal plots are shown in Figure 30. Feeble rural-urban contrast in PM_{2.5} was observed during PMN. No significant gradients in PM_{2.5} were observed in other seasons.

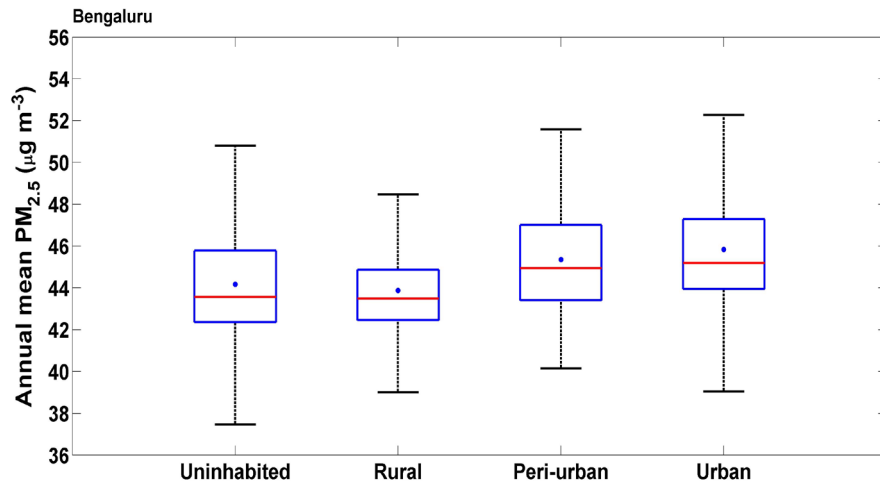


Figure 29: Variations in the annual mean PM_{2.5} of the Bengaluru region for the identified settlement classes

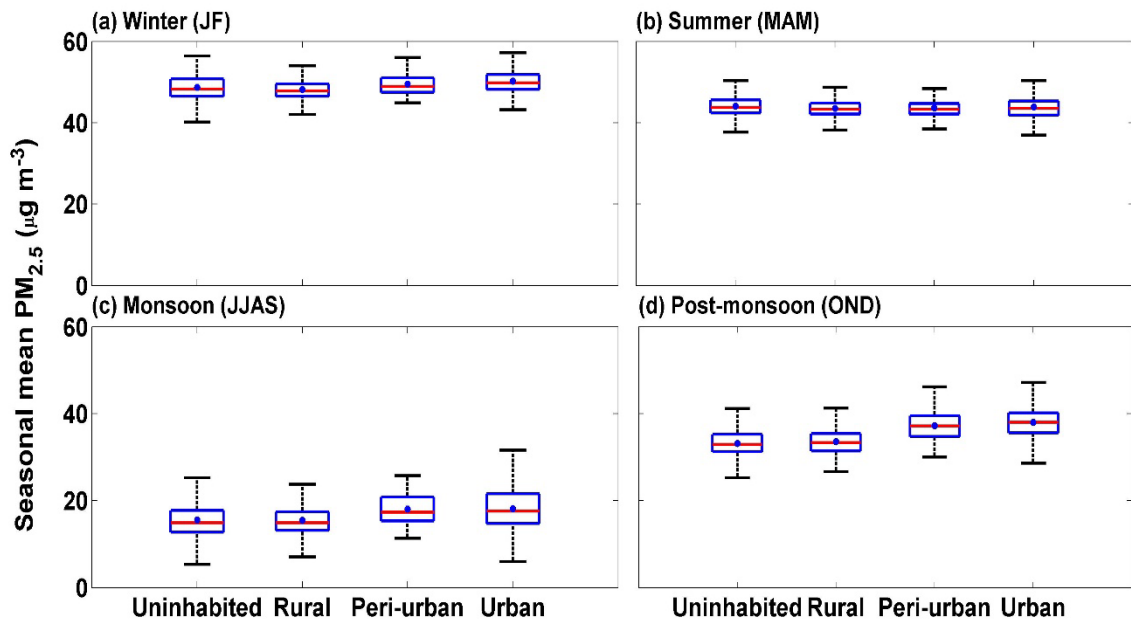


Figure 30: Variations in the seasonal mean PM_{2.5} of the Bengaluru region for the identified settlement classes

Table 8: The annual and season-wise mean PM_{2.5} statistics for different settlement classes of study regions. IQR and SD represent the interquartile range and standard deviation, respectively.

Study region	Settlement class	Season	Mean ($\mu\text{g m}^{-3}$)	SD ($\mu\text{g m}^{-3}$)	Median ($\mu\text{g m}^{-3}$)	IQR ($\mu\text{g m}^{-3}$)
Delhi-NCR	Uninhabited	WIN	132	14.5	130	23.7
		SUM	79	5.5	78	7.2
		MON	56	8.6	57	12.8
		PMN	129	10.3	129	14.8
		Annual	100	6.6	100	9.7
	Rural	WIN	133	15.1	132	24.9
		SUM	79	4.8	78	6.0
		MON	58	7.4	59	10.9
		PMN	131	10.2	130	14.4
		Annual	101	6.5	100	9.8
	Peri-urban	WIN	133	16.2	132	25.3
		SUM	79	3.8	79	5.1
		MON	62	5.3	63	6.6
		PMN	133	12.6	130	18.9
		Annual	102	7.8	100	12.5
	Urban	WIN	151	15.3	153	20.4
		SUM	82	5.0	81	6.8
		MON	60	7.4	61	10.0
		PMN	142	15.3	142	25.1
		Annual	109	9.4	109	14.2
Kanpur	Uninhabited	WIN	139	8.0	138	10.4
		SUM	81	3.1	81	4.3
		MON	71	5.5	71	6.8
		PMN	138	5.9	137	6.8
		Annual	107	3.8	107	4.2
	Rural	WIN	142	8.9	141	10.8
		SUM	82	3.3	81	4.1
		MON	71	5.6	71	7.2
		PMN	138	7.3	137	7.6

	Peri-urban	Annual	108	4.6	107	4.4
		WIN	155	6.5	155	7.2
		SUM	83	4.1	83	5.3
		MON	68	3.6	68	4.6
		PMN	144	5.9	145	5.7
		Annual	112	4.4	113	4.4
	Urban	WIN	154	12.4	154	21.5
		SUM	84	4.2	84	5.9
		MON	69	6.4	69	7.4
		PMN	145	9.6	143	14.3
Annual		114	6.4	113	10.7	
Bengaluru	Uninhabited	WIN	49	2.8	48	4.3
		SUM	44	2.4	44	3.2
		MON	16	4.1	15	5.0
		PMN	33	2.9	33	4.0
		Annual	44	2.4	44	3.4
	Rural	WIN	48	2.4	48	3.0
		SUM	44	2.2	43	2.7
		MON	15	3.4	15	4.2
		PMN	34	3.0	33	4.0
		Annual	44	2.1	43	2.4
	Peri-urban	WIN	50	2.6	49	3.6
		SUM	44	2.2	43	2.5
		MON	18	4.0	17	5.5
		PMN	37	3.6	37	4.8
		Annual	45	2.4	45	3.6
	Urban	WIN	50	3.0	50	3.6
		SUM	44	3.0	44	3.4
		MON	18	4.5	18	6.9
		PMN	38	3.6	38	4.6
		Annual	46	2.9	45	3.3

Annual and seasonal mean PM_{2.5} maps of Indore city (Madhya Pradesh) are given in Appendix B.

8. Conclusions

PM_{2.5} data are critical for urban planning and epidemiological, environmental, and economic studies. However, the limited availability of PM_{2.5} data is a challenge. This study deals with the spatial mapping of high-resolution 1 km × 1 km satellite AOD-based gridded PM_{2.5} estimates and the quantification of PM_{2.5} over select regions in India. The study regions include the urban, peri-urban, and rural areas of Delhi-NCR, Kanpur, and Bengaluru.

In the study, two linear mixed-effects (LME) models were developed to estimate PM_{2.5}, one for the North Indian study regions (Delhi-NCR and Kanpur) and one for the South Indian study region (Bengaluru). A suite of satellite (AOD, CWV, NDVI) and reanalysis meteorological products was considered predictors for these models. The North India model performed better than the South India model in terms of correlation and determination coefficients; their performance is comparable in terms of RMSE and NMRSE.

A key output of this study is high-resolution spatial maps of daily mean PM_{2.5}. Seasonal variations (WIN high and MON low) in PM_{2.5} were observed across the study regions, which are in line with earlier studies. PM_{2.5} in Delhi-NCR and Kanpur region was found to be three times higher than that of the Bengaluru region. Within Delhi-NCR, high PM_{2.5} levels were observed in eastern Delhi-NCT (East, Shahdara, South East, and North East) and other regions (Gurugram, Faridabad, Nuh, Palwal, and Bharatpur). In the Kanpur region, high PM_{2.5} levels were observed in the Kanpur city. In Bengaluru, high levels were observed in the western part of urban Bengaluru and over a few areas in Greater Bengaluru.

The spatial maps generated were used to identify pollution hotspots. The most noticeable PM_{2.5} hotspots were clustered in and around Delhi-NCT. Some isolated hotspots were also observed. Most of the hotspots coincided with urban areas of Delhi-NCR. The seasonal hotspot analysis identified a high PM_{2.5} cluster in the Bharatpur district of Rajasthan, which is mostly a remote region. This could be due to dust storms that are a frequent phenomenon over that region. The Delhi-NCT region—including Shahdara, East, and North East districts—were most severely polluted and identified as hotspots across seasons except MON. Shahdara is one of the oldest and highly urbanised areas of Delhi-NCT, consisting of residential and industrial establishments. MON hotspots need to be considered with caution as the wet deposition (owing to rainfall characteristics) of aerosol particles will not be uniform across the region.

In the Kanpur region, most of the urban zones were identified as PM_{2.5} hotspot areas. Hotspots were identified in most zones (a few zones entirely and a few zones partially) on annual and seasonal scales (except MON). Zones 1 and 4 are relatively small in size, largely unplanned, and are characterised by a very high general residential and slum population density. Hotspots were identified in the grid clusters where industries and residential areas are located. The Panki Power Plant located in Zone 5 could also be a major source of PM_{2.5}. Zones 2 and 6 have a large number of rural settlements along with agricultural areas.

PM_{2.5} hotspots were identified over a lesser spatial extent in the Bengaluru region compared to that of Delhi-NCR and Kanpur regions. Hotspots were identified in Dasarahalli, Rajarajeshwari Nagar, West, South, Bommanahalli, Mahadevapura, and Yelahanka zones. The hotspots coincided with some of the industrial hubs such as Peenya Industrial Area (Dasarahalli zone), ITPL (International Tech Park Ltd.), Graphite India (Mahadevapura zone), and the Karnataka Housing Board industrial area (Yelahanka zone). A recent study by the Karnataka State Pollution Control Board (2020) recorded a Comprehensive Environmental Pollution Index (CEPI) score of ~65.11 for the Peenya Industrial Cluster, declaring it a severely polluted area.

The current study also clearly identified Peenya and surrounding areas as hotspots. The areas include Hegganahalli, Doddanna Industrial Estate, Herohalli, Doddabidarakallu, Nagarabhavi II Stage, and Laggere wards. Bommanahalli also houses red and orange category industries. The Greater Bengaluru PM_{2.5} hotspots (observed during WIN and SUM) coincided with the stone-crushing locations. The noticeable PM_{2.5} hotspot region in Bengaluru rural was located in the Tubagere town, Doddaballapura Taluk. Open burning in and around the town has the potential to enhance PM_{2.5} concentrations.

Based on PM_{2.5} spatial maps and GHMS-SMOD data, variations in PM_{2.5} across several settlement classes (urban, peri-urban, rural, and uninhabited) were studied. For Delhi-NCR and Kanpur regions, a marginal difference in annual PM_{2.5} was observed between rural and urban areas. For the Bengaluru region, no difference was observed between urban, peri-urban, and rural areas, irrespective of seasons. Even though most urban areas were identified as hotspots, no significant difference between urban and rural PM_{2.5} was observed.

Hotspot identification is not only based on the magnitude of PM_{2.5} but also its clustering pattern. One of the potential reasons for the observed shallow gradients is the regional nature of PM_{2.5} (widespread natural and anthropogenic sources). In India, a large portion of the rural population still depends on solid fuels for domestic use, which can contribute significantly to rural PM_{2.5} (Conibear et al., 2018; Balakrishnan et al., 2019). A recent mobile monitoring study in Bengaluru by CSTEP and ILK Labs (2020) observed shallow gradients in on-road PM_{2.5} across urban and peri-urban areas.

8.1 Limitations

The accuracy of the estimated PM_{2.5} depends on several factors:

1. The ground truth data used for training the models were third-party data, which were being collected/monitored by pollution control board authorities. Rigorous quality-check measures were applied while compiling the PM_{2.5} data for the model. However, any unidentified errors/biases in the data can impact the accuracy of the model estimates.
2. India being a tropical country, the frequent presence of clouds can lead to missing data or AOD with higher uncertainty. Only the highest quality AOD (identified based on the quality flag) was used in this study.
3. Missing model PM_{2.5} estimates during MON (due to missing satellite AOD) biased the annual mean values. This bias was higher in the Bengaluru region compared to other study regions.
4. Models in the present study were trained using urban PM_{2.5}; however, predictions/estimates were also made in the peri-urban and rural areas. Given the scale of the study, more peri-urban and rural validation exercises need to be conducted.
5. The spatial density of the monitoring data can also influence the model performance. Compared to Delhi-NCR, other study regions had fewer PM_{2.5} monitoring stations.
6. The model configured in the present study assumed that the temporal (day-to-day) variability in the AOD-PM_{2.5} relationship was spatially invariant. The frequent occurrence of elevated aerosol layers in the northern and western parts of India during SUM months can override this assumption.

9. Way Forward

9.1 Future research recommendations

The economic resources of LMICs do not support establishing a spatially dense network of reference-grade air pollution monitoring stations, as they are expensive. Satellite-AOD- based inexpensive $PM_{2.5}$ estimates were already recommended for developing countries such as India by Brauer et al. (2019) as part of their proposed hybrid monitoring approach. Because of the complexities in satellite-AOD-based $PM_{2.5}$ estimates and based on the experience gained from the current study, we propose the following research recommendations for the future:

1. Given the heterogeneous pollution sources and varied meteorology and topography across the Indian subcontinent, local models are preferred to estimate $PM_{2.5}$ from AOD.
2. Because of the limited regulatory air pollution monitoring data available for training statistical or artificial intelligence models, calibrated $PM_{2.5}$ data from low-cost sensor networks could be used as training data.
3. For accurately locating hotspot areas, a much higher resolution (sub-kilometre) $PM_{2.5}$ is desired. Downscaling methods could be explored to reduce the spatial resolution of the $PM_{2.5}$ estimates further.
4. Statistical methods such as land-use regression (LUR) could be explored in cities that have spatially dense monitoring data (or mobile monitoring data) to overcome the missing $PM_{2.5}$ estimates from satellite models (because of the missing AOD).
5. Suitable auxiliary predictors such as information on aerosol vertical distribution, proxies for elevated aerosols, and cloud optical depth could be explored for improving the accuracy of $PM_{2.5}$ estimates.

9.2 Policy recommendations

Investigations carried out during the study have revealed several interesting spatial characteristics of $PM_{2.5}$. Based on the results obtained, the following policy recommendations are made:

1. Using high-resolution $PM_{2.5}$ maps, policymakers and regulatory authorities could prioritise areas for devising/implementing air pollution mitigation strategies to achieve an effective reduction in the area-averaged $PM_{2.5}$.
2. Continuous ambient air quality monitoring stations could be established in representative rural and peri-urban locations. Data from these stations could aid researchers in fine-tuning their models for achieving more accurate spatial predictions of air pollution levels in non-urban regions.
3. Source apportionment studies could be conducted in representative rural and peri-urban locations to identify and quantify the pollution sources contributing to rural pollution.
4. Though city-specific actions plans are already being implemented in the cities that were studied (Delhi, Kanpur, and Bengaluru), further actions are recommended to regulate polluting activities of unorganised sectors (brick kilns, quarries, etc.) operating in the periphery of these cities.

10. References

- Alduchov, O. A., & Eskridge, R. E. (1996). Improved Magnus form approximation of saturation vapor pressure. *Journal of Applied Meteorology*, 35(4), 601–609. [https://doi.org/10.1175/1520-0450\(1996\)035<0601:IMFAOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2)
- 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>
- Balakrishnan, K., Dey, S., Gupta, T., Dhaliwal, R. S., Brauer, M., Cohen, A. J., ... Sabde, Y. (2019). The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017. *The Lancet Planetary Health*, 3(1), e26–e39. [https://doi.org/10.1016/S2542-5196\(18\)30261-4](https://doi.org/10.1016/S2542-5196(18)30261-4)
- 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>
- Chitranshi, S., Sharma, S. P., Dey, S. (2015). Satellite-based estimates of outdoor particulate pollution (PM₁₀) for Agra City in northern India. *Air Quality, Atmosphere & Health*, 8(1), 55–65. <https://doi.org/10.1007/s11869-014-0271-x>
- Chowdhury, S., Dey, S. (2016). Cause-specific premature death from ambient PM_{2.5} exposure in India: Estimate adjusted for baseline mortality. *Environment International*, 91, 283–290. <https://doi.org/10.1016/j.envint.2016.03.004>
- Chowdhury, S., Dey, S., Di Girolamo, L., Smith, K. R., Pillarisetti, A., Lyapustin, A. (2019). Tracking ambient PM_{2.5} build-up in Delhi national capital region during the dry season over 15 years using a high-resolution (1 km) satellite aerosol dataset. *Atmospheric Environment*, 204, 142–150. <https://doi.org/10.1016/j.atmosenv.2019.02.029>
- Christopher, S., & Gupta, P. (2020). Global distribution of column satellite aerosol optical depth to surface PM_{2.5} relationships. *Remote Sensing*, 12(12), 1985. <https://doi.org/10.3390/rs12121985>
- Conibear, L., Butt, E.W., Knote, C., Arnold, S.R., Spracklen, D.V. (2018). Residential energy use emissions dominate health impacts from exposure to ambient particulate matter in India. *Nature Communications*, 9, 1–9. <https://doi.org/10.1038/s41467-018-02986-7>
- CSTEP & ILK Labs (2020). Mobile-monitoring campaign for air pollution studies in Bengaluru. (CSTEP-WS-2020-04). <https://www.cstep.in/projects-details.php?id=1634>
- Dey, S., Purohit, B., Balyan, P., Dixit, K., Bali, K, M et al. (2020). A Satellite-Based High-Resolution (1-km) Ambient PM_{2.5} Database for India over Two Decades (2000–2019): Applications for Air Quality Management. *Remote Sensing*, 12, 1–22. <https://doi.org/10.3390/rs12233872>
- Dey, S., Di Girolamo, L., van Donkelaar, A., Tripathi, S. N., Gupta, T., Mohan, M. (2012). Variability of outdoor fine particulate (PM_{2.5}) concentration in the Indian Subcontinent: A remote sensing approach. *Remote Sensing of Environment*, 127, 153–161. <https://doi.org/10.1016/j.rse.2012.08.021>

- Duncan, B. N., Prados, A. I., Lamsal, L. N., Liu, Y., Streets, D. G., Gupta, P., ... Burton, S. P. (2014). Satellite data of atmospheric pollution for US air quality applications: Examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid. *Atmospheric Environment*, *94*, 647–662. <https://doi.org/10.1016/j.atmosenv.2014.05.061>
- 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*, *100096*. <https://doi.org/10.1016/j.aeaoa.2020.100096>
- Getis, A., & Ord, J. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24* (3), 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Guo, Y., Tang, Q., Gong, D. Y., Zhang, Z. (2017). Estimating ground-level PM_{2.5} concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sensing of Environment*, *198*, 140–149. <https://doi.org/10.1016/j.rse.2017.06.001>
- Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *Journal of Geophysical Research: Atmospheres*, *114*(D20). <https://doi.org/10.1029/2008JD011496>
- Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y. C., Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, *40*(30), 5880–5892. <https://doi.org/10.1016/j.atmosenv.2006.03.016>
- Guttikunda, S. K., Nishadh, K. A., Gota, S., Singh, P., Chanda, A., Jawahar, P., Asundi, J. (2019). Air quality, emissions, and source contributions analysis for the Greater Bengaluru region of India. *Atmospheric Pollution Research*, *10*(3), 941–953. <https://doi.org/10.1016/j.apr.2019.01.002>
- Habibi, R., Alesheikh, A., Mohammadinia, A., Sharif, M. (2017). An assessment of spatial pattern characterization of air pollution: A case study of CO and PM_{2.5} in Tehran, Iran. *ISPRS International Journal of Geo-Information*, *6*, 270. <https://doi.org/10.3390/ijgi6090270>
- Han, W., Tong, L., Chen, Y., Li, R., Yan, B., Liu, X. (2018). Estimation of high-resolution daily ground-level PM_{2.5} concentration in Beijing 2013–2017 using 1 km MAIAC AOT data. *Applied Sciences*, *8*(12), 2624. <https://doi.org/10.3390/app8122624>
- Health Effects Institute. (2020). State of Global Air 2020. Special Report. Boston, MA: Health Effects Institute. <https://www.stateofglobalair.org/>
- Hoff, R. M., & Christopher, S. A. (2009). Remote sensing of particulate pollution from space: have we reached the promised land? *Journal of the Air & Waste Management Association*, *59*(6), 645–675. <https://doi.org/10.3155/1047-3289.59.6.645>
- Hsu, N. C., Tsay, S. C., King, M. D., Herman, J. R. (2004). Aerosol properties over bright-reflecting source regions. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(3), 557–569. <https://doi.org/10.1109/TGRS.2004.824067>
- Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., ... Liu, Y. (2014). Estimating ground-level PM_{2.5} concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment*, *140*, 220–232. <https://doi.org/10.1016/j.rse.2013.08.032>

- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., ... Liu, Y. (2018). Predicting monthly high-resolution PM_{2.5} concentrations with random forest model in the North China Plain. *Environmental Pollution*, 242, 675–683. <https://doi.org/10.1016/j.envpol.2018.07.016>
- Jethva, H., Torres, O., Yoshida, Y. (2019). Accuracy assessment of MODIS land aerosol optical thickness algorithms using AERONET measurements over North America. *Atmospheric Measurement Techniques*, 12(8). <https://doi.org/10.5194/amt-12-4291-2019>
- Karnataka State Pollution Control Board. (2020). Action plan for severely polluted Peenya Industrial Area/Estate, Bengaluru, Karnataka State. https://kspcb.karnataka.gov.in/sites/default/files/inline-files/4%20peenya_2019_1852019_0.pdf
- Kim, K. H., Kabir, E., Kabir, S. (2015). A review on the human health impact of airborne particulate matter. *Environment International*, 74, 136–143. <https://doi.org/10.1016/j.envint.2014.10.005>
- Kumar, N., Chu, A., Foster, A. (2007). An empirical relationship between PM_{2.5} and aerosol optical depth in Delhi Metropolitan. *Atmospheric Environment*, 41(21), 4492–4503. <https://doi.org/10.1016/j.atmosenv.2007.01.046>
- Lee, H. J., Liu, Y., Coull, B. A., Schwartz, J., Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict PM_{2.5} concentrations. *Atmospheric Chemistry & Physics*, 11(15). <https://doi.org/10.5194/acp-11-7991-2011>
- Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., Eck, T. F. (2010). Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmospheric Chemistry & Physics*, 10(21), 10399. <https://doi.org/10.5194/acp-10-10399-2010>
- Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., Kaufman, Y. J. (2007). Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *Journal of Geophysical Research: Atmospheres*, 112(D13). <https://doi.org/10.1029/2006JD007815>
- Li, Y., Xue, Y., Guang, J., She, L., Fan, C., Chen, G. (2018). Ground-Level PM_{2.5} concentration estimation from satellite data in the Beijing area using a specific particle swarm extinction mass conversion algorithm. *Remote Sensing*, 10(12), 1906. <https://doi.org/10.3390/rs10121906>
- Lyapustin, A. I., Wang, Y., Laszlo, I., Hilker, T., Hall, F. G., Sellers, P. J., ... Korkin, S. V. (2012). Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3. Atmospheric correction. *Remote Sensing of Environment*, 127, 385–393. <https://doi.org/10.1016/j.rse.2012.09.002>
- Lyapustin, A., Wang, Y., Hsu, C., Torres, O., Leptoukh, G., Kalashnikova, O., Korkin, S. (2011). Analysis of MAIAC dust aerosol retrievals from MODIS over North Africa. *Atti della Accademia Peloritana dei Pericolanti-Classe di Scienze Fisiche, Matematiche e Naturali*, 89(S1). <https://doi.org/10.1478/C1V89S1P061>
- Lyapustin, A., Wang, Y., Korkin, S., Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, 11(10). <https://doi.org/10.5194/amt-11-5741-2018>

- Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korokin, S., Remer, L., ... Reid, J. S. (2011). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research: Atmospheres*, 116(D3). <https://doi.org/10.1029/2010JD014985>
- Mandal, S., Madhipatla, K. K., Guttikunda, S., Kloog, I., Prabhakaran, D., Schwartz, J. D., Team, G. H. I. (2020). Ensemble averaging based assessment of spatiotemporal variations in ambient PM_{2.5} concentrations over Delhi, India, during 2010–2016. *Atmospheric Environment*, 224, 117309. <https://doi.org/10.1016/j.atmosenv.2020.117309>
- Martin, R. V., Brauer, M., van Donkelaar, A., Shaddick, G., Narain, U., Dey, S. (2019). No one knows which city has the highest concentration of fine particulate matter. *Atmospheric Environment: X*, 3, 100040. <https://doi.org/10.1016/j.aeaoa.2019.100040>
- Martins, V. S., Lyapustin, A., de Carvalho, L. A., Barbosa, C. C. F., Novo, E. M. L. D. M. (2017). Validation of high-resolution MAIAC aerosol product over South America. *Journal of Geophysical Research: Atmospheres*, 122(14), 7537–7559. <https://doi.org/10.1002/2016JD026301>
- Melchiorri, M., Florczyk, A. J., Freire, S., Schiavina, M., Pesaresi, M., Kemper, T. (2018). Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer. *Remote Sensing*, 10(5), 768. <https://doi.org/10.3390/rs10050768>
- Meng, X., Garay, M. J., Diner, D. J., Kalashnikova, O. V., Xu, J., Liu, Y. (2018). Estimating PM_{2.5} speciation concentrations using prototype 4.4 km-resolution MISR aerosol properties over Southern California. *Atmospheric Environment*, 181, 70–81. <https://doi.org/10.1016/j.atmosenv.2018.03.019>
- Mhawish, A., Banerjee, T., Sorek-Hamer, M., Bilal, M., Lyapustin, A. I., Chatfield, R., Broday, D. M. (2020). Estimation of High-Resolution PM_{2.5} over the Indo-Gangetic Plain by Fusion of Satellite Data, Meteorology, and Land Use Variables. *Environmental Science & Technology*, 54(13), 7891–7900. <https://doi.org/10.1021/acs.est.0c01769>
- Mhawish, A., M., Banerjee, T., Sorek-Hamer, Lyapustin, A., Broday, D. M., Chatfield, R. (2019). Comparison and evaluation of MODIS Multi-angle Implementation of Atmospheric Correction (MAIAC), aerosol product over South Asia. *Remote Sensing of Environment*, 224, 12–28. <https://doi.org/10.1016/j.rse.2019.01.033>
- Ministry of Environment, Forest & Climate Change (MoEFCC), 2019. National Clean Air Programme. http://moef.gov.in/wp-content/uploads/2019/05/NCAP_Report.pdf, accessed on 28 October 2020
- Murray, C. J., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., ... Abegaz, K. H. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258), 1223–1249. [https://doi.org/10.1016/S0140-6736\(20\)30752-2](https://doi.org/10.1016/S0140-6736(20)30752-2)
- 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>
- Schaap, M., Apituley, A., Timmermans, R.M.A., Koelemeijer, R.B.A., Leeuw, G.D. (2009). Exploring the relation between aerosol optical depth and PM_{2.5} at Cabauw, the Netherlands. *Atmospheric Chemistry and Physics*, 9, 909–925. <https://doi.org/10.5194/acp-9-909-2009>

- Schraufnagel, D. E., Balmes, J. R., Cowl, C. T., De Matteis, S., Jung, S. H., Mortimer, K., ... Thurston, G. D. (2019). Air pollution and noncommunicable diseases: A review by the Forum of International Respiratory Societies' Environmental Committee, Part 2: Air pollution and organ systems. *Chest*, *155*(2), 417–426. <https://doi.org/10.1016/j.chest.2018.10.041>
- Song, W., Jia, H., Huang, J., Zhang, Y. (2014). A satellite-based geographically weighted regression model for regional PM_{2.5} estimation over the Pearl River Delta region in China. *Remote Sensing of Environment*, *154*, 1–7. <https://doi.org/10.1016/j.rse.2014.08.008>
- Sreekanth, V. (2013). Satellite derived aerosol optical depth climatology over Bangalore, India. *Advances in Space Research*, *51*(12), 2297–2308. <https://doi.org/10.1016/j.asr.2013.01.022>
- Sreekanth, V., Mahesh, B., Niranjana, K. (2017). Satellite remote sensing of fine particulate air pollutants over Indian mega cities. *Advances in Space Research*, *60*(10), 2268–2276. <https://doi.org/10.1016/j.asr.2017.08.008>
- Superczynski, S. D., Kondragunta, S., Lyapustin, A. I. (2017). Evaluation of the multi-angle implementation of atmospheric correction (MAIAC) aerosol algorithm through intercomparison with VIIRS aerosol products and AERONET. *Journal of Geophysical Research: Atmospheres*, *122*(5), 3005–3022. <https://doi.org/10.1002/2016JD025720>
- van Donkelaar, A., Martin, R. V., Brauer, M., Boys, B. L. (2015). Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environmental Health Perspectives*, *123*(2), 135–143. <https://doi.org/10.1289/ehp.1408646>
- Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21). <https://doi.org/10.1029/2003GL018174>
- Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., Bai, Y. (2015). Daily estimation of ground-level PM_{2.5} concentrations over Beijing using 3 km resolution MODIS AOD. *Environmental Science & Technology*, *49*(20), 12280–12288. <https://doi.org/10.1021/acs.est.5b01413>

11. Appendix A: Supplemental Figures

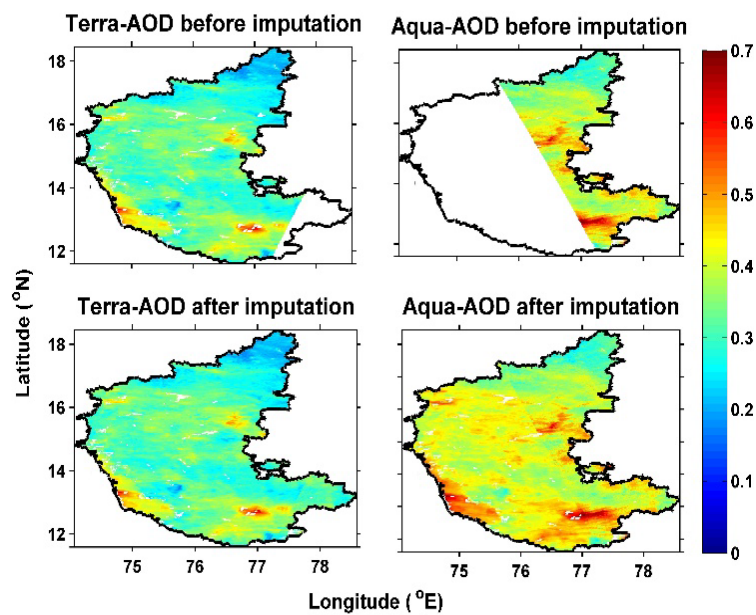


Figure A1: AOD spatial distribution before and after imputation for a typical winter day

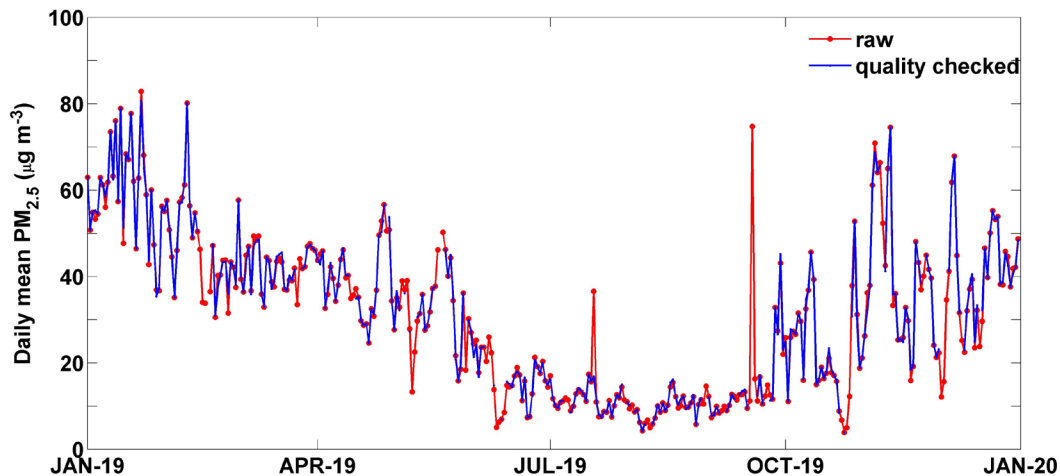


Figure A2: Typical raw and quality-checked daily mean $PM_{2.5}$ time series

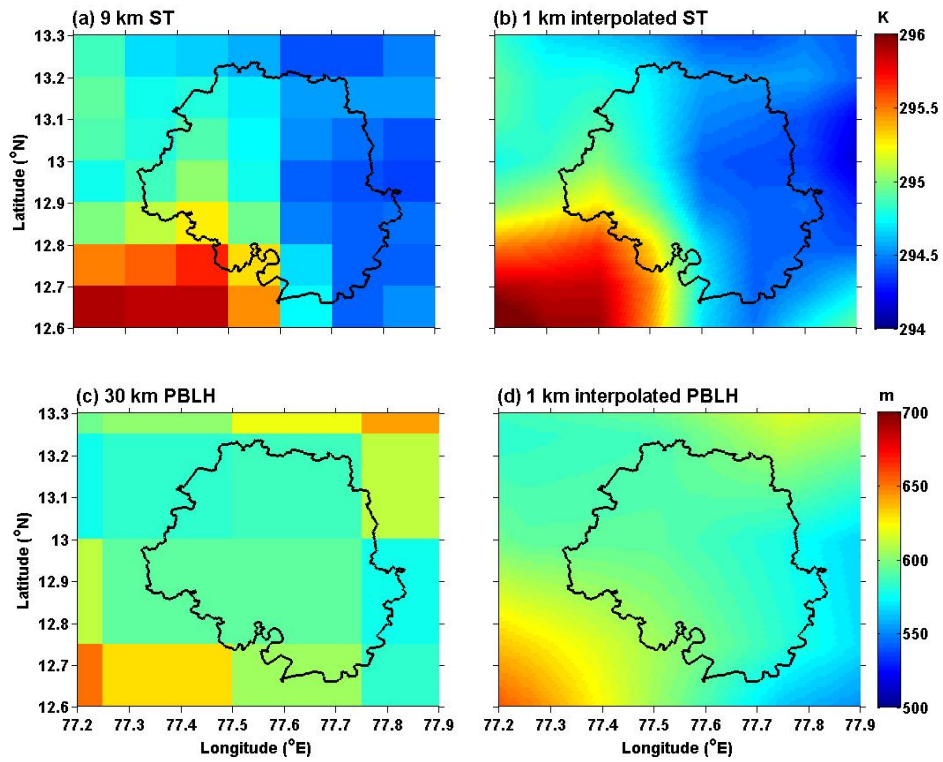


Figure A3: Bilinear interpolation

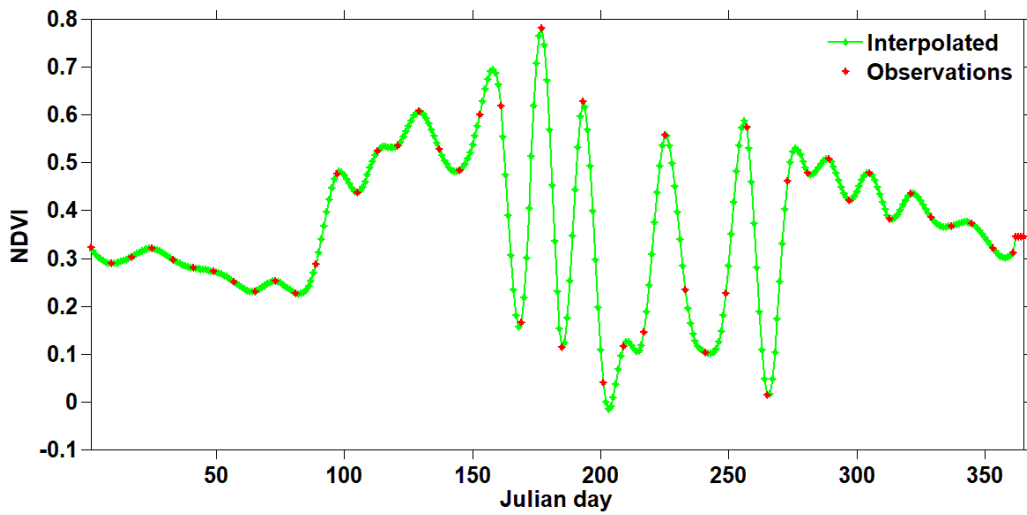


Figure A4: The cubic-spline temporal interpolation of NDVI

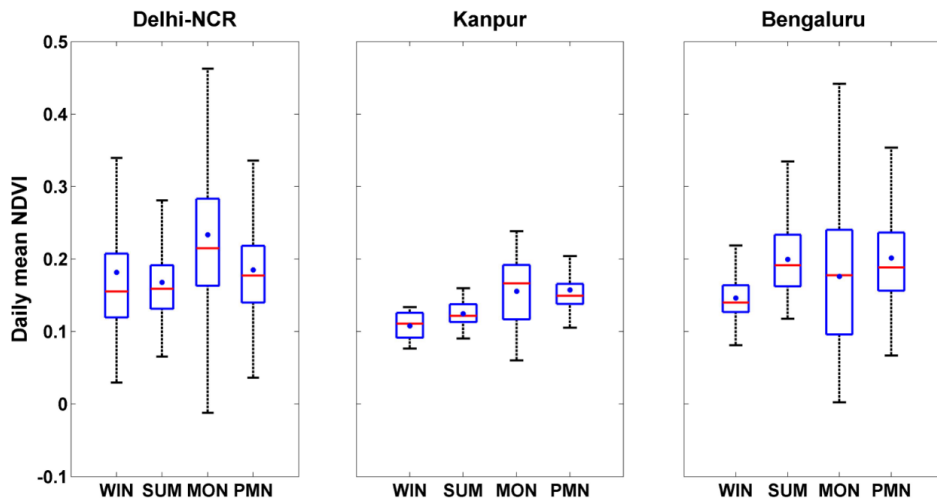


Figure A5: Spatio-temporal variations in the daily mean NDVI

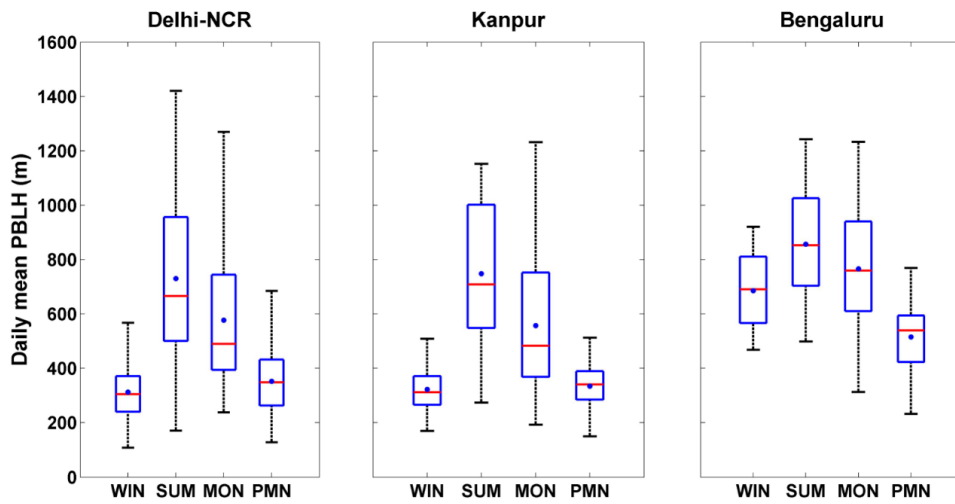


Figure A6: Spatio-temporal variations in the daily mean PBLH

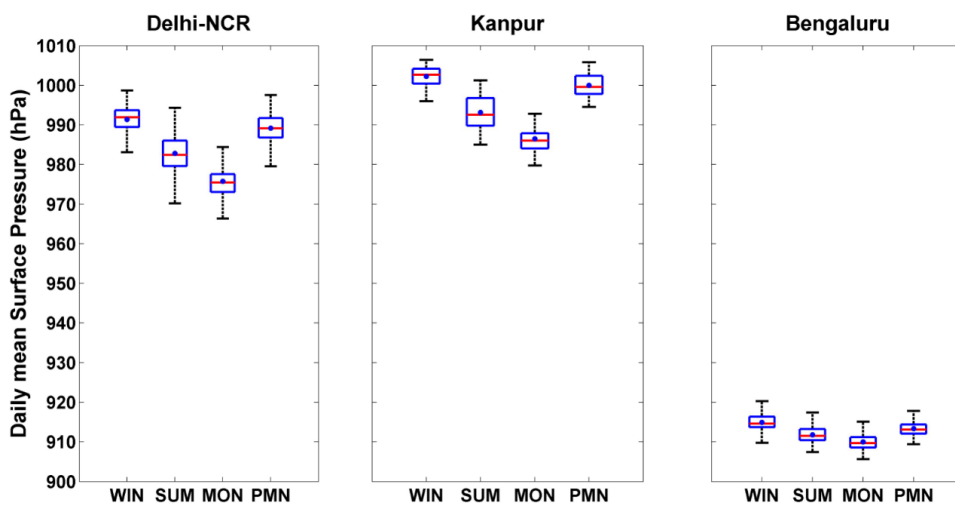


Figure A7: Spatio-temporal variations in the daily mean surface pressure

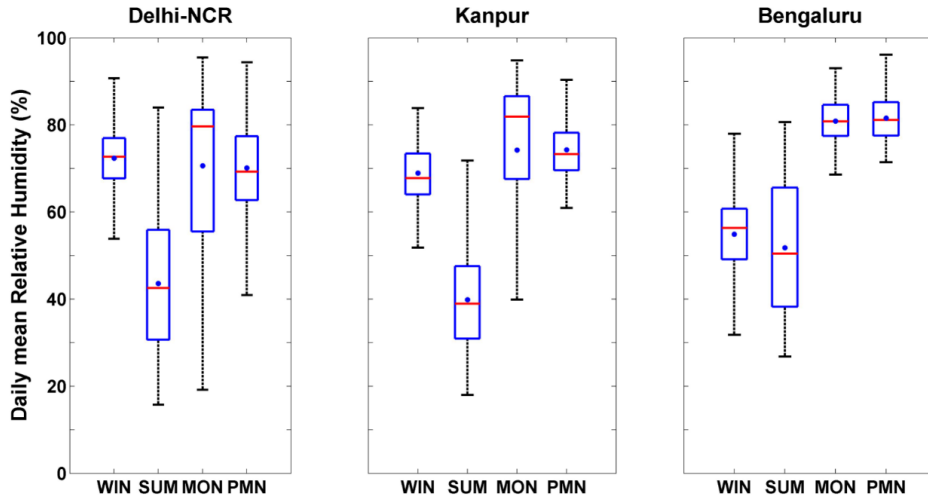


Figure A8: Spatio-temporal variations in the daily mean relative humidity

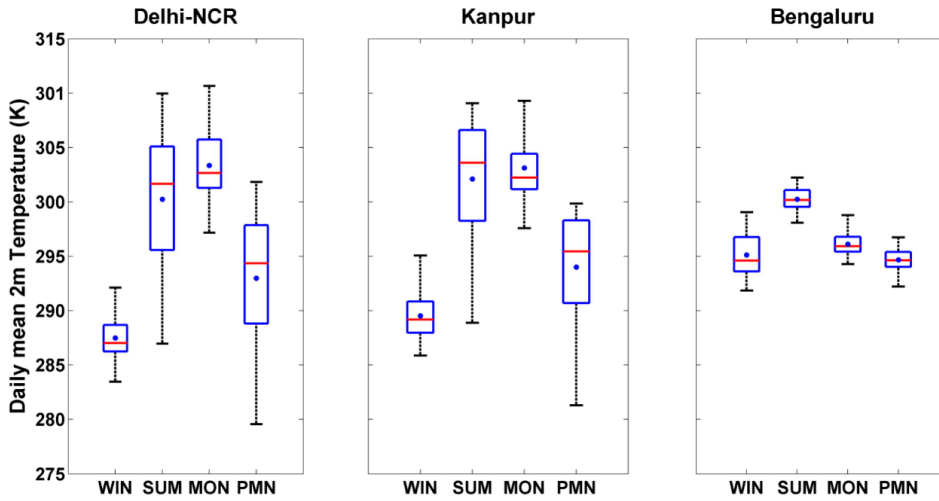


Figure A9: Spatio-temporal variations in the daily mean surface temperature

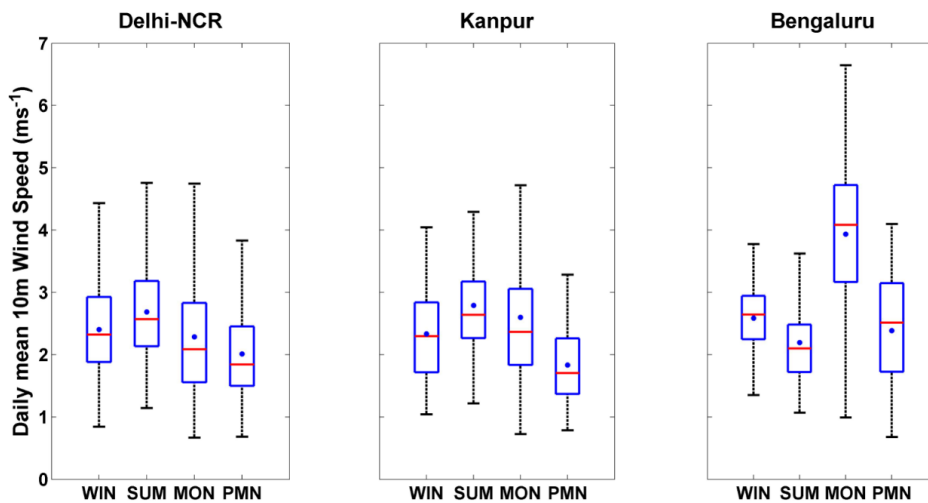


Figure A10: Spatio-temporal variations in the wind speed

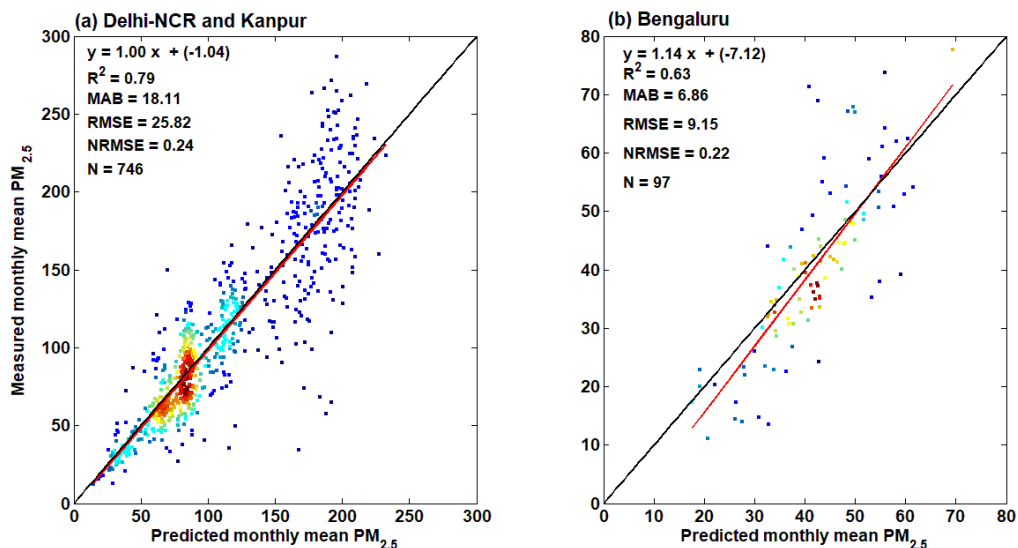


Figure A11: Density scatter plots depicting the model-predicted monthly mean $PM_{2.5}$ and measured monthly mean $PM_{2.5}$. Monthly mean values were derived by aggregating daily mean values to the calendar month.

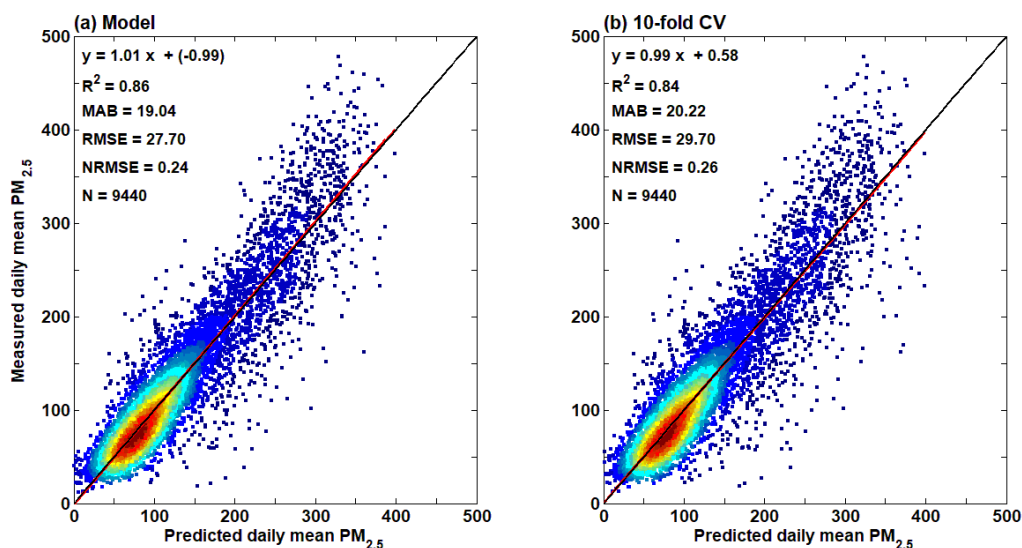


Figure A12: Density scatter plots depicting the model-predicted daily mean $PM_{2.5}$ and measured daily mean $PM_{2.5}$ (the Delhi-NCR and Kanpur model). Here, the LME model was configured with the site-specific random intercept, in addition to the day-specific random slope and intercept.

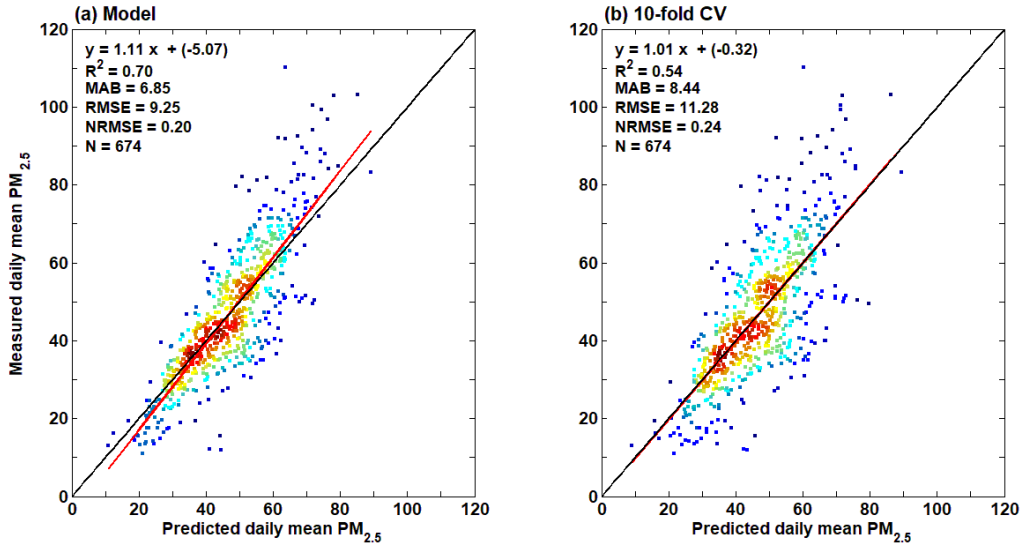


Figure A13: Density scatter plots depicting the model-predicted daily mean PM_{2.5} and measured daily mean PM_{2.5} (the Bengaluru model). Here, the LME model was configured with the site-specific random intercept, in addition to the day-specific random slope and intercept.

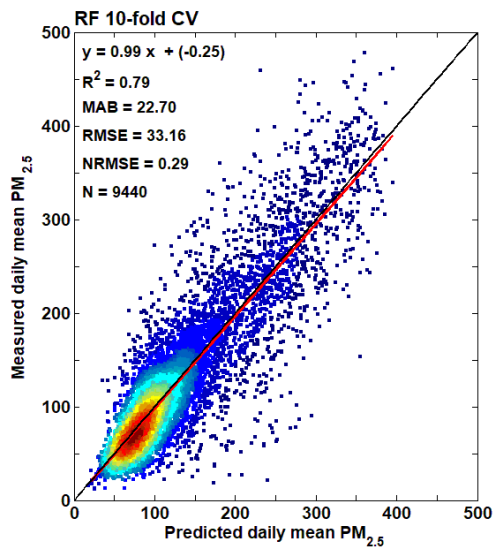


Figure A14: The performance of the random forest model (10-fold CV) for Delhi-NCR and Kanpur

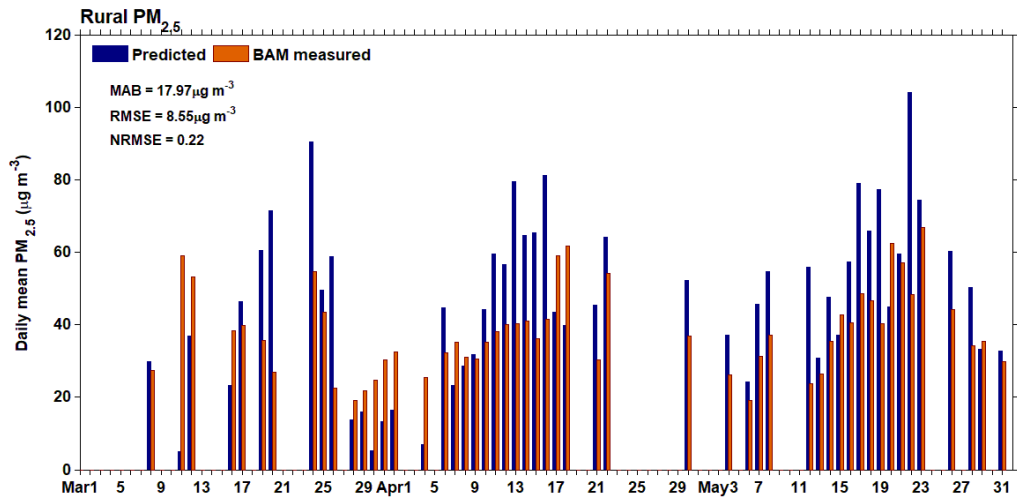


Figure A15: The comparison of the LME-predicted and measured PM_{2.5} for a rural location in UP

Table A1: Details of the CAAQMS stations from which the PM_{2.5} data are used in this study along with PM_{2.5} and MODIS-AOD annual statistics. N indicates the number of daily mean data points available in the study year. SD represents the standard deviation.

State	City	Continuous Ambient Air Quality Station (CAAQMS)	Latitude (° N)	Longitude (° E)	PM _{2.5} (µg m ⁻³)		AOD	
					N	Mean ± SD	N	Mean ± SD
Karnataka	Bengaluru	Peenya, Bengaluru - CPCB	13.0270	77.4941	304	37 ± 16.9	104	0.45 ± 0.19
Karnataka	Bengaluru	BTM Layout, Bengaluru - CPCB	12.9135	77.5951	319	40 ± 23.1	48	0.41 ± 0.19
Karnataka	Bengaluru	BWSSB Kadabesanahalli, Bengaluru - CPCB	12.9352	77.6814	279	42 ± 19.5	49	0.37 ± 0.18
Karnataka	Bengaluru	Bapuji Nagar, Bengaluru - KSPCB	12.9519	77.5398	313	38 ± 13.9	50	0.47 ± 0.20
Karnataka	Bengaluru	Hebbal, Bengaluru - KSPCB	13.0292	77.5859	312	32 ± 17.6	59	0.37 ± 0.15
Karnataka	Bengaluru	Hombegowda Nagar, Bengaluru - KSPCB	12.9385	77.5901	292	28 ± 15.3	72	0.39 ± 0.17
Karnataka	Bengaluru	Jayanagar 5th Block, Bengaluru - KSPCB	12.9210	77.5849	290	31 ± 17.7	75	0.39 ± 0.18
Karnataka	Bengaluru	Silk Board, Bengaluru - KSPCB	12.9173	77.6228	293	32 ± 13.8	59	0.34 ± 0.16
Karnataka	Chikkaballapur	Chikkaballapur Rural, Chikkaballapur - KSPCB	13.4288	77.7314	252	30 ± 15.0	88	0.42 ± 0.18
Karnataka	Hubballi	Deshpande Nagar, Hubballi - KSPCB	15.3518	75.1407	184	35 ± 18.9	106	0.38 ± 0.13
Karnataka	Kalaburagi	Lal Bahadur Shastri Nagar, Kalaburagi - KSPCB	17.3220	76.8226	282	37 ± 24.0	64	0.43 ± 0.12
Karnataka	Mysuru	Hebbal 1st Stage, Mysuru - KSPCB	12.2104	76.3738	115	24 ± 09.2	93	0.36 ± 0.13
Karnataka	Ramanagara	Vijay Nagar, Ramanagara - KSPCB	12.7334	77.2981	86	24 ± 10.9	94	0.41 ± 0.15
Karnataka	Yadgir	Collector Office, Yadgir - KSPCB	16.7602	77.1428	93	38 ± 18.7	95	0.46 ± 0.16
Delhi	Delhi	Alipur, Delhi - DPCC	28.8153	77.1530	318	108 ± 79.3	165	0.66 ± 0.50
Delhi	Delhi	Anand Vihar, Delhi - DPCC	28.6468	77.3160	335	125 ± 97.6	172	0.69 ± 0.52
Delhi	Delhi	Ashok Vihar, Delhi - DPCC	28.6954	77.1817	351	117 ± 95.7	144	0.67 ± 0.53

Delhi	Delhi	Aya Nagar, Delhi - IMD	28.4707	77.1099	323	80 ± 62.4	165	0.71 ± 0.58
Delhi	Delhi	Bawana, Delhi - DPCC	28.7762	77.0511	342	130 ± 91.8	167	0.62 ± 0.47
Delhi	Delhi	Burari Crossing, Delhi - IMD	28.7257	77.2012	266	106 ± 72.7	138	0.68 ± 0.56
Delhi	Delhi	CRRRI Mathura Road, Delhi - IMD	28.5512	77.2736	347	100 ± 79.5	159	0.65 ± 0.52
Delhi	Delhi	DTU, Delhi - CPCB	28.7500	77.1113	340	116 ± 83.8	152	0.62 ± 0.45
Delhi	Delhi	Dr Karni Singh Shooting Range, Delhi - DPCC	28.4986	77.2648	343	96 ± 80.5	153	0.71 ± 0.55
Delhi	Delhi	Dwarka Sector 8, Delhi - DPCC	28.5710	77.0719	355	106 ± 80.1	159	0.70 ± 0.66
Delhi	Delhi	IGI Airport (T3), Delhi - IMD	28.5628	77.1180	312	91 ± 72.6	163	0.64 ± 0.54
Delhi	Delhi	IHBAS, Dilshad Garden, Delhi - CPCB	28.6812	77.3025	343	114 ± 71.1	182	0.70 ± 0.54
Delhi	Delhi	ITO, Delhi - CPCB	28.6286	77.2411	339	108 ± 77.8	166	0.71 ± 0.56
Delhi	Delhi	Jahangirpuri, Delhi - DPCC	28.7328	77.1706	347	126 ± 91.3	145	0.66 ± 0.53
Delhi	Delhi	Jawaharlal Nehru Stadium, Delhi - DPCC	28.5803	77.2338	342	102 ± 88.6	177	0.69 ± 0.53
Delhi	Delhi	Lodhi Road, Delhi - IMD	28.5918	77.2273	325	83 ± 63.3	175	0.68 ± 0.53
Delhi	Delhi	Major Dhyan Chand National Stadium, Delhi - DPCC	28.6113	77.2377	350	91 ± 69.3	168	0.72 ± 0.54
Delhi	Delhi	Mandir Marg, Delhi - DPCC	28.6364	77.2011	343	102 ± 74.2	160	0.72 ± 0.58
Delhi	Delhi	Mundka, Delhi - DPCC	28.6847	77.0766	343	124 ± 96.6	152	0.64 ± 0.51
Delhi	Delhi	NSIT Dwarka, Delhi - CPCB	28.6091	77.0325	354	110 ± 61.2	155	0.66 ± 0.55
Delhi	Delhi	Najafgarh, Delhi - DPCC	28.5702	76.9338	324	93 ± 66.2	171	0.66 ± 0.54
Delhi	Delhi	Narela, Delhi - DPCC	28.8228	77.1020	348	105 ± 75.2	168	0.66 ± 0.51
Delhi	Delhi	Nehru Nagar, Delhi - DPCC	28.5679	77.2505	356	119 ± 104.7	151	0.65 ± 0.49
Delhi	Delhi	North Campus, DU, Delhi - IMD	28.6574	77.1585	288	106 ± 82.6	156	0.64 ± 0.50

Delhi	Delhi	Okhla Phase-2, Delhi - DPCC	28.5308	77.2713	341	106 ± 90.4	141	0.67 ± 0.55
Delhi	Delhi	Patparganj, Delhi - DPCC	28.6237	77.2872	349	95 ± 76.7	164	0.67 ± 0.48
Delhi	Delhi	Punjabi Bagh, Delhi - DPCC	28.6740	77.1310	340	104 ± 83.5	149	0.65 ± 0.56
Delhi	Delhi	Pusa DPCC	28.6396	77.1463	347	98 ± 81.6	154	0.75 ± 0.57
Delhi	Delhi	Pusa, Delhi - IMD	28.6396	77.1463	350	79 ± 61.2	154	0.75 ± 0.57
Delhi	Delhi	R K Puram, Delhi - DPCC	28.5633	77.1869	239	109 ± 85.0	157	0.72 ± 0.59
Delhi	Delhi	Rohini, Delhi - DPCC	28.7325	77.1199	347	128 ± 97.6	140	0.66 ± 0.54
Delhi	Delhi	Shadipur, Delhi - CPCB	28.6515	77.1473	344	116 ± 68.7	153	0.64 ± 0.50
Delhi	Delhi	Sirifort, Delhi - CPCB	28.5504	77.2159	342	107 ± 82.5	154	0.69 ± 0.56
Delhi	Delhi	Sonia Vihar, Delhi - DPCC	28.7105	77.2495	337	104 ± 77.2	161	0.65 ± 0.48
Delhi	Delhi	Sri Aurobindo Marg, Delhi - DPCC	28.5313	77.1902	353	88 ± 71.0	157	0.60 ± 0.59
Delhi	Delhi	Vivek Vihar, Delhi - DPCC	28.6723	77.3153	341	111 ± 90.6	177	0.72 ± 0.57
Delhi	Delhi	Wazirpur, Delhi - DPCC	28.6998	77.1655	349	133 ± 101.1	146	0.66 ± 0.53
Delhi	Delhi	US Diplomatic Post: New Delhi	28.6358	77.2245	298	103 ± 82.4	172	0.71 ± 0.57
UP	Agra	Sanjay Palace, Agra - UPPCB	27.1987	78.0060	336	71 ± 42.7	156	0.60 ± 0.49
UP	Baghpat	New Collectorate, Baghpat - UPPCB	28.9748	77.2134	346	98 ± 67.5	161	0.59 ± 0.39
UP	Bulandshahr	Yamunapuram, Bulandshahr - UPPCB	28.4070	77.8498	274	92 ± 63.9	167	0.62 ± 0.50
UP	Ghaziabad	Vasundhara, Ghaziabad - UPPCB	28.6603	77.3573	333	119 ± 98.4	175	0.68 ± 0.49
UP	Ghaziabad	Indirapuram, Ghaziabad - UPPCB	28.6462	77.3581	269	109 ± 90.4	174	0.68 ± 0.50
UP	Ghaziabad	Loni, Ghaziabad - UPPCB	28.7573	77.2788	254	118 ± 98.2	164	0.63 ± 0.49
UP	Ghaziabad	Sanjay Nagar, Ghaziabad - UPPCB	28.6854	77.4538	273	113 ± 86.8	170	0.66 ± 0.50

UP	Greater Noida	Knowledge Park - III, Greater Noida - UPPCB	28.4727	77.4820	336	102 ± 82.1	161	0.70 ± 0.58
UP	Greater Noida	Knowledge Park - V, Greater Noida - UPPCB	28.5571	77.4537	182	109 ± 102.9	169	0.67 ± 0.52
UP	Hapur	Anand Vihar, Hapur - UPPCB	28.7256	77.7497	240	92 ± 66.8	175	0.63 ± 0.46
UP	Kanpur	Nehru Nagar, Kanpur - UPPCB	26.4703	80.3230	333	93 ± 78.3	142	0.61 ± 0.59
UP	Lucknow	Lalbagh, Lucknow - CPCB	26.8459	80.9366	348	92 ± 61.7	164	0.56 ± 0.40
UP	Lucknow	Talkatora District Industries Center, Lucknow - CPCB	26.8340	80.8917	349	119 ± 67.5	167	0.60 ± 0.43
UP	Lucknow	Central School, Lucknow - CPCB	26.8821	80.9303	338	78 ± 52.5	160	0.56 ± 0.41
UP	Lucknow	Nishant Ganj, Lucknow - UPPCB	26.8714	80.9571	164	97 ± 65.2	163	0.56 ± 0.40
UP	Lucknow	Gomti Nagar, Lucknow - UPPCB	26.8681	81.0051	103	94 ± 58.5	164	0.56 ± 0.39
UP	Meerut	Jai Bhim Nagar, Meerut - UPPCB	28.9536	77.7623	101	121 ± 71.3	165	0.58 ± 0.37
UP	Meerut	Pallavpuram Phase 2, Meerut - UPPCB	29.0635	77.7097	68	158 ± 65.8	171	0.56 ± 0.42
UP	Meerut	Ganga Nagar, Meerut - UPPCB	28.9993	77.7590	80	139 ± 63.6	175	0.56 ± 0.36
UP	Moradabad	Lajpat Nagar, Moradabad - UPPCB	28.8253	78.7213	239	108 ± 68.7	165	0.53 ± 0.36
UP	Muzaffarnagar	New Mandi, Muzaffarnagar - UPPCB	29.4724	77.7194	255	88 ± 55.5	170	0.51 ± 0.33
UP	Noida	Sector - 62, Noida - IMD	28.6245	77.3577	353	102 ± 82.5	162	0.67 ± 0.50
UP	Noida	Sector - 125, Noida - UPPCB	28.5448	77.3231	324	115 ± 87.5	125	0.73 ± 0.59
UP	Noida	Sector-1, Noida - UPPCB	28.5898	77.3101	176	111 ± 114.7	149	0.69 ± 0.59
UP	Noida	Sector-116, Noida - UPPCB	28.5692	77.3938	154	123 ± 111.8	167	0.66 ± 0.48
UP	Varanasi	Ardhali Bazar, Varanasi - UPPCB	25.3506	82.9083	303	97 ± 50.7	173	0.58 ± 0.35
Haryana	Rohtak	MD University, Rohtak - HSPCB	28.5212	76.3714	350	77 ± 49.7	105	0.51 ± 0.33
Haryana	Faridabad	Sector- 16A, Faridabad - HSPCB	28.4088	77.3099	302	97 ± 80.2	95	0.59 ± 0.44

Haryana	Manesar	Sector-2 IMT, Manesar - HSPCB	28.3607	76.9361	342	84 ± 45.2	76	0.51 ± 0.44
Haryana	Panipat	Sector-18, Panipat - HSPCB	29.4980	76.9933	316	71 ± 57.4	109	0.54 ± 0.35
Haryana	Palwal	Shyam Nagar, Palwal - HSPCB	28.1486	77.3321	335	90 ± 53.7	104	0.56 ± 0.45
Haryana	Gurugram	Vikas Sadan, Gurugram - HSPCB	28.4501	77.0263	323	89 ± 57.9	100	0.58 ± 0.55
Haryana	Bahadurgarh	Arya Nagar, Bahadurgarh - HSPCB	28.6701	76.9254	343	70 ± 48.1	102	0.66 ± 0.65
Haryana	Dharuhera	Municipal Corporation Office, Dharuhera - HSPCB	28.2068	76.7997	319	75 ± 42.0	83	0.47 ± 0.33
Haryana	Ballabgarh	Nathu Colony, Ballabgarh - HSPCB	28.3419	77.3197	318	85 ± 53.1	78	0.53 ± 0.43
Rajasthan	Bhiwadi	RIICO Ind. Area III, Bhiwadi - RSPCB	28.1949	76.8623	336	103 ± 45.4	101	0.47 ± 0.35
Haryana	Sonipat	Murthal, Sonipat - HSPCB	29.0272	77.0621	301	65 ± 36.3	123	0.54 ± 0.35

12. Appendix B: Satellite AOD–Derived PM_{2.5} Maps of Indore City

Indore is the most populous and the largest city in the state of Madhya Pradesh. The city is distributed over a land area of 530 km² and has a population of around 1.9 million (2011 census). Indore is surrounded by many major industrial areas, namely Pithampur, the Sanwer industrial belt, and Laxmibainagar Industrial Area. Currently, Indore has only one installed CAAQMS.

The study period was from December 2018 to May 2020. The daily mean PM_{2.5} data from all the CAAQMSs installed across Madhya Pradesh were used to train the LME model (Equation 4). The list of CAAQMSs along with their geographical coordinates, the yearly mean (for the year 2019) PM_{2.5}, and AOD over those station locations are given in Table B1.

Table B1: CAAQMSs from which PM_{2.5} data is used for building the Indore model. SD and N represent the standard deviation and the number of daily mean data points.

City	CAAQMS	Latitude (° N)	Longitude (° E)	PM _{2.5} (µg m ⁻³)		AOD	
				N	Mean (SD)	N	Mean (SD)
Bhopal	TT Nagar	23.2336	77.4006	103	67 (32.7)	139	0.35 (0.14)
Dewas	Bhopal Chauraha	22.9683	76.0641	340	42 (20.4)	145	0.40 (0.17)
Indore	Chhoti Gwaltoli	22.7196	75.8577	102	58 (26.0)	154	0.41 (0.18)
Jabalpur	Marhatal	23.1686	79.9322	101	75 (37.0)	166	0.41 (0.18)
Katni	Gole Bazar	23.8343	80.3894	98	95 (42.9)	154	0.42 (0.21)
Mandideep	Industrial Area	23.1084	77.5114	345	41 (22.6)	156	0.42 (0.21)
Pithampur	Industrial Area	22.6248	75.6752	354	40 (21.6)	154	0.41 (0.17)
Ratlam	Shastri Nagar	23.3317	75.0460	270	47 (18.7)	134	0.37 (0.16)
Singrauli	Vindhyachal STPS	24.1090	82.6456	236	87 (45.6)	133	0.45 (0.25)
Ujjain	Mahakaleswar Temple	23.1827	75.7682	338	43 (24.8)	133	0.43 (0.21)
Domoh	Shrivastav Colony	23.8175	79.4462	310	42 (26.3)	137	0.41 (0.23)
Gwalior	City Centre	26.2034	78.1933	No data in 2019		151	0.52 (0.40)
Gwalior	Phool Bagh, Gwalior - Mondelez Ind. Food	26.2105	78.1710	No data in 2019		159	0.52 (0.41)
Sagar	Deen Dayal Nagar	23.8640	78.8029	No data in 2019		146	0.36 (0.17)

Figure B1 shows the model performance in terms of 10-fold CV and LOOCV. Model R² (0.73) and NRMSE (0.29) values were slightly higher than the other two models in the study.

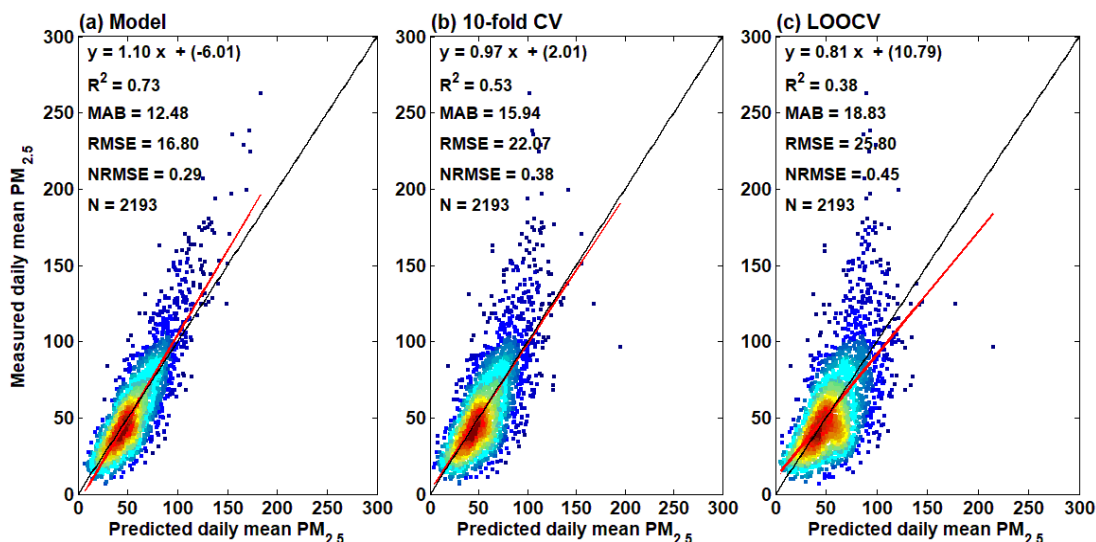


Figure B1: Density scatter plots depicting the model-predicted daily mean $PM_{2.5}$ and measured daily mean $PM_{2.5}$ (the Indore model).

Figure B2 shows the annual mean $PM_{2.5}$ maps for the calendar year 2019 over Indore and surrounding regions. The annual mean $PM_{2.5}$ concentrations varied between 42 and $58 \mu g m^{-3}$ with maximum concentrations ($\sim 58 \mu g m^{-3}$) observed over northern parts of Indore. The spatially averaged annual mean $PM_{2.5}$ values were $\sim 51 \mu g m^{-3}$. The frequency distribution of the gridded spatial $PM_{2.5}$ values is also shown in Figure B2. Indore and surrounding regions include a total of 799 one km grids. Similar to other study regions, Indore’s annual mean $PM_{2.5}$ over all the spatial grids exceeded the Indian national annual standard ($40 \mu g m^{-3}$).

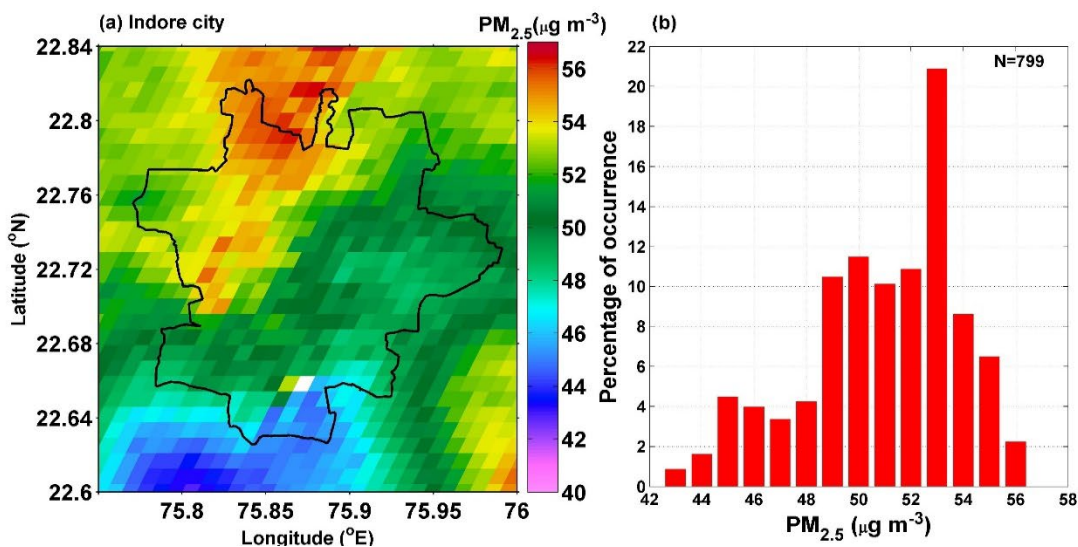


Figure B2: The spatial map of the annual mean $PM_{2.5}$ over Indore for the year 2019 and surrounding regions and their frequency distribution

Figure B3 depicts the seasonal mean maps of $PM_{2.5}$. Similar to Delhi-NCR, Kanpur, and Bengaluru, Indore’s seasonal maps also exhibited significant spatial and temporal variations in $PM_{2.5}$. Seasonally, PMN was the most polluted, followed by WIN and SUM. MON was characterised by the lowest $PM_{2.5}$ levels. In SUM 2020, predicted $PM_{2.5}$ levels were much lower compared to SUM 2019 over all the grids. This was because of the nationwide COVID-19 lockdowns.

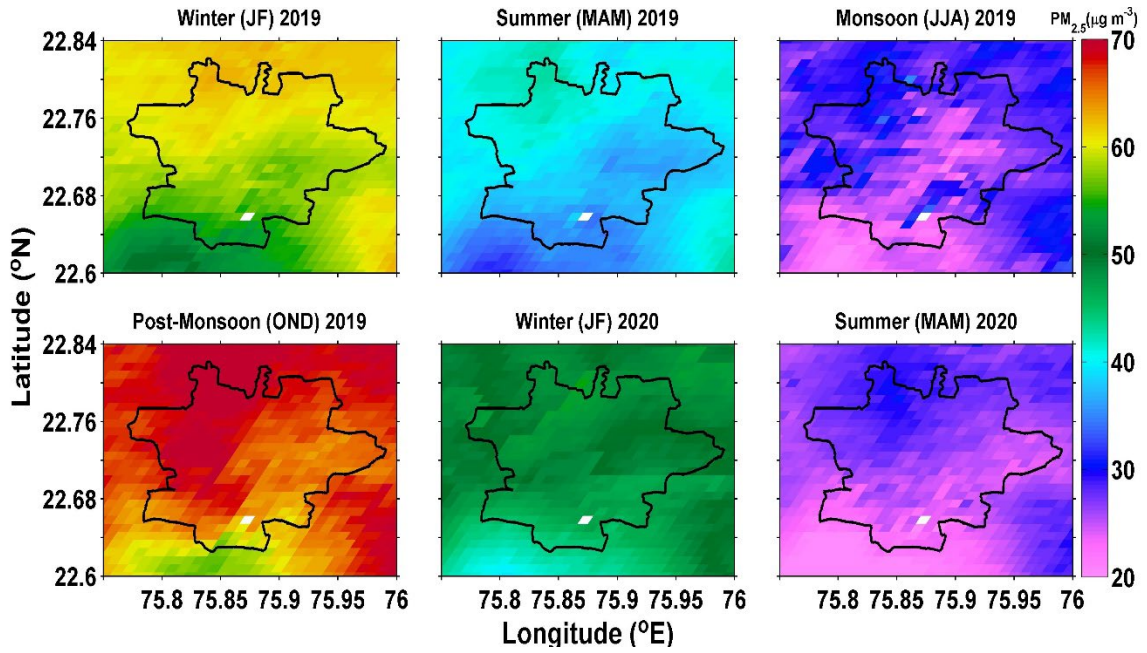


Figure B3: The seasonal mean maps of PM_{2.5} over Indore. The solid black line represents the Indore municipal area

Table B2: The seasonal mean PM_{2.5} statistics for Indore. SD represents the standard deviation.

Indore	Mean (SD) ($\mu\text{g m}^{-3}$)
WIN 2019	59 (3.1)
SUM 2019	39 (2.4)
MON 2019	27 (2.7)
PMN 2019	66 (3.4)
WIN 2020	49 (3.2)
SUM 2020	26 (2.7)



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](https://twitter.com/cstep_india)