

Second stage Project Report
On
**COVID Lockdown Impact on Air Quality:
Satellite and Ground-Based Measurements**

Submitted in the partial fulfilment of the requirement for
the Degree of M. Tech. in Environmental Science and Engineering
by
Madhumitha S
(Roll Number:213180014)

Under the supervision of
Prof. Abhishek Chakraborty



**Environmental Science and Engineering Department
INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY**

June 16, 2023

Approval sheet

This seminar report entitled "**COVID Lockdown Impact on Air Quality: Satellite and Ground-Based Measurements**" prepared by Madhumitha S (Roll No. 213180014), is hereby approved for submission.

Digital Signature
Abhishek Chakraborty (10001846)
15-Jun-23 08:30:12 PM

Prof. Abhishek Chakraborty

(Supervisor)

Date: 16-06-23

Place: Mumbai

Declaration

I **Madhumitha S (Roll No. 213180014)** understand that plagiarism is defined as any one or combination of the following:

1. Uncredited verbatim copying of individual sentences, paragraphs, or illustrations (such as graphs, diagrams, etc.).
2. Uncredited improper paraphrasing of pages or paragraphs (by changing a few words or phrases or rearranging the original sentence order)
3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did or wrote what.

I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

I affirm that no portion of my work can be considered as plagiarized, and I take full responsibility if such a complaint occurs. I understand fully well that the adviser/guide/co-guide of the seminar report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Full Signature of the Student with Date:



Name of the Student: Madhumitha S

Roll Number: 213180011

Date: 16-06-2023

Acknowledgements

I am filled with deep appreciation and gratitude for the love and support of my mother, sister, and fiancé. My sister's financial contributions towards my post-graduation have left me immensely grateful. Additionally, I extend my heartfelt thanks to my friends Bharrathi, Aarthi, Ilamathy, Nandhini, Ayan, Rushikesh, and Keshmina for their support throughout my project and placements. I am grateful to my friends at IIT Bombay, Muthu, Srinithi, Baisakhi, Progga, and Shanta, for the memorable times we spent together, whether it was going out for food or exploring new places, which created a sense of belonging and trust. I extend my gratitude to my IPT teammates Sachin, Harish, Harshit, Pranjal, and Uddesh for their assistance in handling challenging situations on the ground and organizing IPT events.

I am indebted to my guide, Professor Abhishek, for his exceptional patience and motivation during difficult times. The invaluable assistance and constant availability of Prof. Abhishek and co-guide Dr. Karn Vohra in addressing technical queries deserve my sincere gratitude. I also want to express my gratitude to Dr. Karn Vohra for providing tremendous support during my placement preparation. The enthusiasm displayed by my guides has improved my quality of work. I dedicate my success in placements and MTP project to both of my guides, as their support made these achievements possible. I would like to acknowledge Professor Suparna Mukherji for instilling in me a passion for research while working under her during my seminar project and TAship. Furthermore, I thank Professor Sanjeev Chaudhari for providing emotional support and helping me take care of my mental health during difficult situations. I would also like to express my appreciation to my fellow lab mates, particularly Mrs. Shruti and Mr. Debayan, for their assistance in refining my interpretations and their moral support throughout this journey. I am grateful to my lab mates Sathish, Senthil and Krishnakant for creating a better and joyful lab environment.

Abstract

Air Pollution is a serious threat to the nation's health and has to be addressed with stringent pollution control strategies and laws. The COVID-19 nationwide lockdown has given us an eccentric opportunity to study the impact of anthropogenic emissions switch off situations on air quality on a global and regional scale. In this study, we investigate the impact of COVID-19 lockdown on PM_{2.5} and NO₂ and other criteria air pollutants over the Mumbai Metropolitan Region (MMR) using in-situ ambient air quality measurements provided by Centre Pollution Control Board (CPCB) and other freely available satellite and satellite derived products. Our study also reveals potential limitations of using Continuous ambient air quality monitoring station (CAAQMS) data with respect to data quality issues. We developed the first publicly available automated tool to obtain quality-assured air quality datasets by identifying and addressing these data issues. We find that after data-cleaning, 15 out of 40 sites become non-compliant with National Ambient Air Quality Standards (NAAQS), and there is a significant increase of 11.4% in the Air Quality Index and a 20% decrease in the NO_x-to-NO₂ ratio on average across all sites. These have implications for the global research community, health burden assessments, and policy formulation. Data quality issues were exacerbated during the COVID-19 lockdown and worsened in 2021. Our study underscores the urgent need for ratified CAAMQS data in light of serious data quality issues. Post data cleaning, we used CAAQMS data to recalibrate satellite estimated PM_{2.5} data. Using a random forest (RF) model trained on historical PM_{2.5} data and weather variables, we predicted business-as-usual (BAU) conditions during the COVID years of 2020 and 2021. The ground level PM_{2.5} was found to be in good agreement with satellite derived values and could reproduce the variability in ground station measurements. Our analysis reveals a significant increase in PM_{2.5} levels from the early 2000s until 2009, followed by a stabilization of pollution levels. We found that wind directions from

the east, south, and southeast contributed to MMR pollution, while winds from the southwest helped reduce it. During the COVID-19 pandemic, reduced industrial activity and decreased vehicular emissions led to lower PM_{2.5} levels. However, as the MMR transitioned into the post-COVID phase, pollution levels gradually increased with a PM_{2.5} increase of more than 200% compared to the expected rise in BAU 2021. These findings highlight the vulnerability of the MMR to pollution rebounds in the absence of long-term and sustained pollution abatement measures.

Keywords: Lockdown, CPCB, Air quality, Satellite measurements, PM_{2.5}, NO₂

Contributions

Madhumitha S, M. Tech 2nd year student, ESED Department: Carried out all the statistical analysis, visualizations, and interpretations and drafted the manuscript.

Dr. Abhishek Chakraborty, Assistant Professor at Indian Institute of Technology, Bombay & **Dr Karn Vohra**, Research Fellow, UCL: guided throughout the project with their ideation and contributed to thesis writing and proofreading.

Bharrathi A S, M. Tech 2nd year student, ESED Department: Contributed to data retrieval from CAAQMS dashboard and data visualizations in data quality assessment studies (Chapter 4).

Note: Chapter 4 is under revision in ES&T letters and hence is not considered plagiarism.

Urgent Issues regarding Real-time Air Quality Monitoring Data in India: Unveiling Solutions and Implications

Madhumitha S^{1†}, Karn Vohra^{2‡*}, Abhishek Chakraborty^{1*}, and Bharrathi A S¹

¹Environmental Science and Engineering Department, Indian Institute of Technology Bombay, Mumbai, Maharashtra, India.

²Department of Geography, University College London, London, United Kingdom.

†Contributed equally to this work

*Correspondence to: Abhishek Chakraborty (abhishekc@iitb.ac.in) and Karn Vohra (k.vohra@ucl.ac.uk)

Contents

Approval sheet.....	ii
Declaration.....	iii
Acknowledgements	iv
Abstract.....	v
Contribution	vii
List of Figures.....	x
List of Tables	xvi
Abbreviations	xvii
Chapter 1: Introduction	1
1.1 The Threat of Air Pollution: Impact on Health	1
1.2 Air Pollution Challenges and Climate Change Implications in Mumbai.....	1
1.3 COVID Lockdown Impact in Air Quality.....	2
2.1 Objectives of the Report.....	5
Chapter 2: Methodology.....	6
2.1 Study Area.....	6
2.2 Mumbai Metropolitan region shape file:.....	8
2.3. Air Quality Data	8
2.3.1 Ground station measurements:	8
2.3.2 Satellite measurements	9
2.3.3 Meteorological data	11
2.4 Methodology and Analysis	11
2.4.1 Methodology for Comparing Satellite data with CPCB Data	12
2.4.2 Methodology for estimating wind speed from reanalysis data:.....	12
2.4.3 Statistical Analysis	13
2.4.4 Business as Usual analysis	14
Chapter 3: Addressing the CAAQMS Data Quality Issues	16
3.1 Necessity of assessing the data quality	16
3.2 Methodology of AirPy data cleaning tool	18
3.2.1. Pollutant data	18
3.2.2. Data quality control	18
3.3 Data quality issues and implications of cleaning	26
3.3.1 Uncovering the persistent data quality issues in CAAQMS data.....	26
3.3.2 Implications of addressing the air quality data issues	34
3.3.3 Limitations of our data cleaning tool.....	42

3.4.4. Implications for future advancements in data quality management	44
3.4.4. Evaluating the Impact of Data Cleaning on Air Quality Studies	44
3.4.5. Code availability.....	47
Chapter 4: Preliminary Exploratory Data Analysis.....	48
4.1 Satellite Based Ground Level Estimates:	48
4.1.1 Spatial Variation in Particulate Matter Response due to Lockdown:.....	48
4.1.2 Decrease in NO ₂ during Lockdown & NO ₂ Hotspots in MMR:	52
4.1.3 Increase in Tropospheric O ₃ during Lockdown.....	57
4.1.4 Reduction in SO ₂ surface concentration estimates during Lockdown	58
4.2 Comparison of Satellite estimates with ground-based measurements	59
4.3 Long term trends in ground station measurements	62
Chapter 5: Validation of satellite-based ground level PM_{2.5} estimation.....	68
5.1 Methodology:	68
5.1.1 Hybrid PM _{2.5} estimates.....	68
5.1.2. Validation between PM _{2.5} satellite estimates and ground-based records	69
5.1.3. Model performance before recalibration of Satellite measurements.....	69
5.1.4. Concentration dependent bias in Satellite estimates.....	72
5.1.3. Model performance after recalibration of Satellite measurements.....	74
Chapter 6: The Impact of COVID-19 & Post-COVID Trends on Air Quality in MMR	75
6.1 Predicting Business-as-Usual PM _{2.5} Levels for COVID Years	75
6.1.1 Business-as-Usual (BAU) Model development for COVID Years.....	75
6.1.2 Feature importance in Business-as-Usual RF model.....	76
6.2 Weather and PM _{2.5} trend over past two decades.	78
6.3 Monthly variation in PM _{2.5} during lockdown	86
Chapter 7: Conclusions	92
Chapter 8: Annexure	94
Reference	96

List of Figures

Chapter 2: Methodology.....6

Figure 2.1. CAAQMS stations available in MMR Region. The districts within MMR region are given in different colors in panel (a) and the annotations in the panel (b) denote the CAAQMS station names.....7

Figure 2.2. Georeferencing of MMR. The toposheet is projected on the OSM standard map with EPSG: 32643 -WGS 84 projection (b). Post georeferencing, the polygon was digitalized over it and exported as shapefile (.shp).....8

Figure 2.3. Methodology for Business as Usual prediction for 2020 and 2021.....14

Chapter 3: Addressing the CAAQMS Data Quality Issues16

Figure 3.1: Spatial coverage of Continuous Ambient Air Quality Monitoring Stations (CAAQMS) in India. We study 40 sites operational in 2019-2021 (red) and rest of the sites are shown in blue. Insets show the zoomed-in maps for Delhi (1), Mumbai (2), and Kolkata (3). State boundaries are from GitHub repositories (Josh Brobst, 2020; Sajjad Anwar, 2014).....19

Figure 3.2. Determining the optimal threshold for the consecutive repeat algorithm. The performance metrics, including accuracy score, precision score, F1 score, and recall score, are plotted against the y-axis on the left for different thresholds in blue color, while the false positive rate is plotted against the y-axis on the right for different thresholds in red color.....20

Figure 3.3. Major data quality concerns in CAAQMS data in India. All panels show a time series of 15-minute records. Raw data in all panels are shown in blue. Consecutive repeats are in orange in panel (a) and outliers occurring at the same time each day are in green in panel (b). Panel (c) shows the inconsistency in reporting units. Data records after 21st August 2020 have a mismatch between the reported NO_x (blue) and NO_x calculated from NO and NO₂ (red). The inset in panel (c) shows a scatter plot with two populations of data highlighting this mismatch. Data with consistent reporting units is in blue, inconsistent in red and 1:1 line is in black.33

Figure 3.4. Monthly trends in the percentage of data flagged as consecutive repeats and outliers across 2019-2021 for all six pollutants. The percentages of data categorised as consecutive repeats (panels a-c) and outliers (panels d-f) are calculated by considering the total available data as the denominator and normalised across months and sites. The box plots in panels (g) and (h) illustrate the monthly trends in the percentage of consecutive repeats and outliers for all three years, with the box color indicating the year. Each data point in the box plot represents the normalised percentages for each site in each month across all three years.28

Figure 3.5. Correlation between Mobility and Consecutive Repeats of Air Quality Data during Lockdown Phases in India. The plot presents a correlation analysis between mobility and the consecutive repeats of air quality data during the 2020 lockdown phases in India. The graph shows the average mobility in blue on the x-axis and the normalized percentages of consecutive repeats of air quality data in red on the y-axis and the different lockdown phases (phase 1 to phase 4) are marked by green planes.29

Figure 3.6. Identification of data with no diurnal variability using our algorithm. The primary panel presents a time series of a 15-minute O₃ dataset from a monitoring site in

western India. The main panels highlight the repetitive fluctuations removed by the tool in red and its corresponding diurnal variability in inset (a). The other data are shown in blue in main panel with the valid diurnal variation in inset (b)30

Figure 3.7: Identification of Extreme Outliers in Pollutant Data. Each data point corresponds to the 15-minute average, with extreme outliers marked in red.....31

Figure 3.8: Comparison of NO, NO₂, and NO_x before and after addressing unit inconsistency issues. The figure displays the correct NO, NO₂, and NO_x data in blue, which comply with the CPCB protocol and are reported in appropriate units, and were unchanged post unit correction as it populates along 1:1 black line. The incorrect reporting of NO, NO₂, and NO_x in ppb is shown in red, with the incorrect reporting of NO_x in $\mu\text{g m}^{-3}$ shown in violet, and incorrect reporting of NO in ppb shown in green.34

Figure 3.9: Impact of addressing unit inconsistency issue on diurnal variability of NO₂ concentrations. Panels (a) and (c) show the NO_x and NO+NO₂ data before and after addressing the unit inconsistency issue in NO and NO₂, respectively, for the month of August 2020 at a monitoring site in northern India. Panels (b) and (d) show the corresponding diurnal variability of the NO₂ data associated with panels (a) and (c), respectively. The diurnal pattern of NO₂ concentrations before addressing the unit inconsistency issue (panel a) is shown in red, and after the issue was addressed (panel d), it is also shown in red.35

Figure 3.10. Impact of air quality data cleaning on annual means of O₃ (a), NO₂ (b), PM_{2.5} (c) and PM₁₀ (d) in 2020. The sites that showed changes in compliance with the NAAQS are indicated with a black outline. Size of data points represents the annual mean concentrations of air pollutants. The map of India shows state boundaries and the insets show district boundaries, with insets 1 for Mumbai, 2 for Kolkata, and 3 for Delhi. The annual mean concentrations were calculated by averaging the data at a 24-hour rolling window for PM_{2.5}, PM₁₀, NO₂, and at an 8-hour rolling window for O₃. Grey data points indicate sites with less than 75% of data available post-cleaning. The state and district boundaries are from GitHub repositories (Josh Brobst, 2020; Sajjad Anwar, 2014).....36

Figure 3.11: Impact of data cleaning on annual mean of O₃ (a), NO₂ (b), PM_{2.5} (c), PM₁₀ (d) and change in the compliance with NAAQS standards for 2019. The size of the data points in each panel is proportional to the annual mean of NO₂ after the data cleaning process, and black outlines indicate sites where there were changes in compliance with NAAQS standards. The India map shows state boundaries and the insets show district boundaries, with inset 1 for Mumbai, 2 for Kolkata, and 3 for Delhi. Grey data points indicate sites with less than 75% data availability before or after data cleaning. The state and district boundaries are from GitHub repositories (Josh Brobst, 2020; Sajjad Anwar, 2014).38

Figure 3.12: Impact of data cleaning on annual mean of O₃ (a), NO₂ (b), PM_{2.5} (c), PM₁₀ (d) and change in the compliance with NAAQS standards for 2021. The size of the data points in each panel is proportional to the annual mean of NO₂ after the data cleaning process, and black outlines indicate sites where there were changes in compliance with NAAQS standards. The India map shows state boundaries and the insets show district boundaries, with inset 1 for Mumbai, 2 for Kolkata, and 3 for Delhi. Grey data points indicate sites with less than 75% data availability before or after data-cleaning. The state and district boundaries are from GitHub repositories (Josh Brobst, 2020; Sajjad Anwar, 2014).39

Figure 3.13. Impact of data cleaning on air quality index and NO_x to NO₂ ratio. The Figure illustrates the impact of data cleaning on the ratio of NO_x to NO₂ (in panel a) and the air quality index (in panel b) for the year 2020. The shape colors distinguish between the

presence or absence of unit inconsistency issues in the oxides of nitrogen data and 1:1 line is in black.41

Figure 3.14. Identification of dubious waves in trace gas datasets. Panels a and c display the time series data for O₃ and NO₂ respectively. Panels b and d exhibit the corresponding diurnal variability for these datasets.42

Figure 3.15. Identification of dubious data spread and partial cleaning using a data cleaning tool. Panels (a) and (c) show the NO data from a monitoring site in northern India that was identified and removed (depicted in red) using the tool. Panels (b) and (d) display the corresponding diurnal variation of NO for the years 2019 and 2021, respectively. Each data point represents a 15-minute average. A comparison of the 2019 and 2021 data suggests that the 2019 data may contain invalid observations.43

Figure 3.16. Unusual limits on PM dataset. Panel (a) displays the PM_{2.5} data with an upper limit of 59 $\mu\text{g m}^{-3}$ in red, and panel (b) shows the minimum possible value of PM₁₀ constrained at 10 $\mu\text{g m}^{-3}$ in red and other observations are given in blue. Each datapoint represents 15-minute averages.44

Figure 3.17. Impact of NO₂ data cleaning in published air quality studies in India. The figure illustrates the absolute NO₂ ($\mu\text{g m}^{-3}$) values reported in air quality studies(Jain & Sharma, 2020; P. Kumari & Toshniwal, 2020a; Mandal et al., 2021; M. Sarkar et al., 2021; Shukla et al., 2022; Siddiqui et al., 2022) using CAAQMS from Delhi, Mumbai, and Kolkata. The x-axis represents the absolute NO₂ values reported in the studies, while the y-axis represents the averages computed after data cleaning in the current study. The size of each data point is proportional to the change in NO₂ values observed after data cleaning, relative to the absolute values reported by the CAAQMS sites.46

Figure 3.18. . Percentage of invalid data and its influence on annual mean pollutant levels (2019-2021): Dubious data identification within CPCB and Non-CPCB sites CAAQMS data. Panels (a-b) display the percentage of consecutive repeats (a) and outliers (b) removed from the total available data for each site from 2019 to 2021. Panels (c-e) illustrate the relative change in annual mean after the removal of consecutive repeats (c), outliers (d), and ensuring unit consistency (e). Bar colours distinguish data from different authorities: CPCB in blue, non-CPCB (SPCBs, IITM, IMD) in orange, with each data point inside the bar representing a specific site in each year (2019-2020).47

Chapter 4: Preliminary Exploratory Data Analysis.....48

Figure 4.1: Monthly mean PM_{2.5} ($\mu\text{g/m}^3$) in $0.01^\circ \times 0.01^\circ$ for year 2019 and 2020 (Jan – Jun). The arrows in 2020 visualizations indicates the increase or decrease in PM_{2.5} in 2020 with respect to 2019 (Red color arrow – indicating the increase). Annotations inside each subplots has descriptive statistics such a minimum, maximum, standard deviation, mean concentration for values within MMR50

Figure 4.2. April month PM_{2.5} ($\mu\text{g/m}^3$) in $0.01^\circ \times 0.01^\circ$ for year 2019 and 2020 and the absolute increase in PM_{2.5} in 2019, the outskirts of MMR with respect to 2020. Annotations inside each subplot have descriptive statistics such a minimum, maximum, standard deviation, and mean concentration for values within MMR.....51

Figure 4.3. Land cover composition of MMR region in 2020 and 2019. MMR central hotspot is covered with barren land whereas Mumbai central hotspot is composed of built-up area.53

Figure 4.4: Monthly mean NO₂ (ppb) in $0.01^\circ \times 0.01^\circ$ for the years 2019 and 2020 (Jan – Jun). Concentration of NO₂ is in ppb. Annotations inside each subplot have descriptive

statistics such a minimum, maximum, standard deviation, mean concentration for values within MMR.....	54
Figure 4.5: Overview on impact of wind speed on NO ₂ hotspots. Panel 1 data has NO ₂ satellite estimates and Panel two indicates the wind speed and direction. Calm wind condition is denoted by circles and wind barbs denote the direction of wind (For simplicity, pink arrow denotes the wind direction).	55
Figure 4.6: Response of NO ₂ and PM _{2.5} satellite estimates during 2020 (COVID year) and 2019 (March – Pre COVID month; April, May – COVID Month). Annotations: ns: p <= 1; *: 1e-2 < p <= 5e-2; **: 1e-3 < p <= 1e-2; ***: 1e-4 < p <= 1e-3; ****: p <= 1e-4.	56
Figure 4.7: April O ₃ Data comparison using Satellite measurement: OMI/Aura O ₃ (O ₃) DOAS Total Column Level 3 data 0.25° x 0.25°. Concentration in DU (Total column). Source: Giovanni.....	57
Figure 4.8: Time Averaged Map of SO ₂ Surface Mass Concentration (ENSEMBLE) for April months 0.5°x 0.625°. [MERRA-2 Model M2TMNXAER v5.12.4] Concentration in kg m ⁻³	58
Figure 4.9: Satellite estimations (After MPCB data correction using temporal correction factor) and surface measurements for	59
Figure 4.11: Comparison PM ₁₀ during lockdown month (April). Pre lockdown years (Prl) 2017 - 2019, Lockdown (L) – 2020, Post Lockdown (PL) – 2021 -2022. (Concentration of PM ₁₀ is in $\mu\text{g}/\text{m}^3$). Annotations: ns: p <= 1; *: 1e-2 < p <= 5e-2; **: 1e-3 < p <= 1e-2; ***: 1e-4 < p <= 1e-3; ****: p <= 1e-4.	63
Figure 4.12: Comparison NO ₂ during lockdown month (April). Pre lockdown years (Prl) 2017 - 2019, Lockdown – 2020, Post Lockdown – 2021 -2022. (Concentration of NO ₂ is in $\mu\text{g}/\text{m}^3$). Annotations: ns: p <= 1; *: 1e-2 < p <= 5e-2; **: 1e-3 < p <= 1e-2; ***: 1e-4 < p <= 1e-3; ****: p <= 1e-4.	64
Figure 4.13: Comparison O ₃ during lockdown month (April). Pre lockdown years (Prl) 2017 - 2019, Lockdown – 2020, Post Lockdown – 2021 -2022. (Concentration of NO ₂ is in $\mu\text{g}/\text{m}^3$). Annotations: ns: p <= 1; *: 1e-2 < p <= 5e-2; **: 1e-3 < p <= 1e-2; ***: 1e-4 < p <= 1e-3; ****: p <= 1e-4.	65
Figure 4.14: Comparison.....	66
Figure 4.15: Comparison of O ₃ from 2019 to 2021. Pre lockdown year - 2019, Lockdown – 2020 (Concentration of O ₃ is in $\mu\text{g}/\text{m}^3$). Annotations on the marker denote the data availability used to calculate the yearly mean. N denotes the total number of stations with data available.....	67
Chapter 5: Validation of satellite-based ground level PM_{2.5} estimation.....	68
Figure 5.1: Comparison of PM _{2.5, sat} and PM _{2.5, GM} from 2019-2021. Each data point in the scatterplot corresponds to monthly averages at each individual district. The PM _{2.5, GM} of each district was average of one or more station functional between 2019-2021. The color of each scatter point denotes the amount of data available at the time of calculation of monthly average at each CPCB site which is again aggregated at district levels.	70
Figure 5.2: Analysis of PM _{2.5, sat} Discrepancy: Ground-Based Measurements vs. Satellite Estimates (2019-2021). Panel (a) illustrates the dependency of error bias on the concentration of PM _{2.5, GM} . Panel (b) explores the independence of the bias on monthly variability, while panel (c) investigates its independence on annual variability.	71

Figure 5.4: Comparison of PM_{2.5}, Adj sat and PM_{2.5}, GM from 2019-202. This figure illustrates the comparison between PM_{2.5} estimates obtained from the adjusted satellite data (PM_{2.5}, Adj sat) and ground-based measurements (PM_{2.5}, GM) for the period spanning 2019 to 2021. Each data point in the scatterplot represents the monthly averages for individual districts. The PM_{2.5}, GM values for each district were calculated as the average of one or more functional monitoring stations operating between 2019 and 2021. The color coding of each scatter point indicates the availability of data during the calculation of monthly averages at each CPCB site, which were then aggregated at the district level. Notably, only the PM_{2.5} estimates from the Vasai and Thane districts were used to train the model for predicting PM_{2.5}, GM values for all three years. Furthermore, predictions for the years 1994 to 2020 were also generated based on this trained model73

Figure 5.3: Improvement in agreement of PM_{2.5}, sat with PM_{2.5}, GM post adjustment: Ground-Based Measurements vs. Satellite Estimates (2019-2021). Panel a illustrates the dependency of error bias on the concentration of PM_{2.5}, GM. Panel b explores the independence of the bias on monthly variability, while panel c investigates its independence on annual variability. 74

Chapter 6: The Impact of COVID & Post-COVID Trends on Air Quality in MMR75

Figure 6.1. Variable importance and partial dependency plot for RF BAU model. Panel (a) shows the importance of features in the model, whereas panel (b) shows the effect of each feature in the model's performance.77

Figure 6.2. Trends in PM_{2.5} between 1994-2021 using the after smoothening the pollutant data from MMR region. The shading in (a) & (b) shows the estimated 95 % confidence intervals. The solid red line shows Theil Sen, and the dashed red lines show the 95 % confidence intervals for the trend based on resampling methods for 2019-2021 in observed PM_{2.5} in (d) and predicted PM_{2.5} in (c). * Indicates the significance of the trend.78

Figure 6.3. Trends in PM_{2.5} at MMR region from 1994-2021. The plot shows the deseasonalised monthly mean concentrations of PM_{2.5}. The solid red line shows the trend estimate by TheilSen and the dashed red lines show the 95 % confidence intervals for the trend based on resampling methods. For example, the overall trend from Southeast direction is shown at the top-left as 1.77 (μgm^{-3}) per year and the 95 % confidence intervals in the slope from 1.5–1.98 μgm^{-3} /year. The *** show that the trend is significant to the 0.001 level.81

Figure 6.4. Trends in PM_{2.5} at MMR region from 2016-2021. The plot shows the deseasonalised monthly mean concentrations of PM_{2.5}. The solid red line shows the trend estimate by TheilSen and the dashed red lines show the 95 % confidence intervals for the trend based on resampling methods. For example, the overall trend from the Southeast direction is shown at the top-left as 1.77 μgm^{-3} per year and the 95 % confidence intervals in the slope from 1.5–1.98 μgm^{-3} /year. The *** show that the trend is significant to the 0.001 level.82

Figure 6.5. PolarPlot of PM_{2.5} concentrations at MMR based on the CPF function for a range of percentile intervals from 0–25, 25–50, 50–75, 75–100 (left to right) for 2019 and 2020.....83

Figure 6.6: PolarPlot function for the mean concentration of PM_{2.5}.....85

Figure 6.7. Analysis on wind speed/direction frequencies by year (2016-2021). Wind speeds are split into intervals shown by the scale in each panel. The grey circles show the 10-50 % frequencies.86

Figure 6.8: Trend in PM _{2.5} pollution across districts of MMR. Panel b shows the observed and BAU model output. Panel c shows the increase in observed PM _{2.5} in post-COVID year (2021)	88
Figure 6.9: Comparison between actual PM _{2.5} and BAU Model value for 2017-2021 at monthly.....	90
Figure 6.10: Trend in PM _{2.5} pollution across districts of MMR between 2020-2019. Boxes compare the average monthly mean with observed in green and predicted BAU in red. The black solid line shows 0% relative change between the observed and model and relative change for each month is given in blue solid line.	91

List of Tables

Table 1: COVID-19 Lockdown in India.....	3
Table 2.1. CPCB Monitoring sites and Parameters being reported. Pre-COVID studies were only possible for two stations such as Bandra Mumbai MPCB and Pimpleshwar Mandir, Thane – MPCB, since air quality data for 2019 is completely available only for these sites. Source: CPCB.....	10
Table 3.1. Categories of unit inconsistency identified across all sites during 2019, 2020 and 2021. The central tendency for $(NO_2 + NO)/NO_x$ ratio for a particular category is given along with the superscript and subscript which indicates the empirically determined upper and lower limit to categorize the data into four categories.	23
Table 3.2. Unit inconsistency for oxides of nitrogen across sites in India during 2019, 2020, and 2021. The table displays sites reporting oxides of nitrogen in accordance with the CPCB protocol (green) and those that do not comply (C2 in red). Sites dubiously reporting in more than one unit during the year are in yellow. NA refers to no data for the year.....	24
Table 4.1: Wind speed in MMR region from 2018-2019.....	60
Table 4.2: Sites available for long term trend studies in Maharashtra.....	60
Table 8.1. Parameters measured by CAAQMS stations in MMR region.....	94

Abbreviations

MMR	Mumbai Metropolitan Region
MMR	Mumbai Metropolitan Region District administration
PM _{2.5}	Particulate Matter < 2.5 μm Aerodynamic diameter
PM ₁₀	Particulate Matter < 10 μm Aerodynamic diameter
NO	Nitrous Oxide
NO _x	Nitrogen Oxides
NO ₂	Nitrogen dioxide
O ₃	Ozone
SO ₂	Sulphur dioxide
VOC	Volatile Organic Compounds
MODIS	Moderate Resolution Imaging Spectroradiometer
TROPOMI	Tropospheric Monitoring Instrument
OMI	Ozone Monitoring Instrument
LANDSAT	Land Remote-Sensing Satellite
GEOS	Goddard Earth Observing System
AQI	Air Quality Index
COVID-19	Coronavirus Disease 2019
CPCB	Central Pollution Control Board
SPCB	State Pollution Control Board
MPCB	Maharashtra Pollution Control Board
CAAQMS	Continuous Ambient Air Quality Monitoring Stations
MERRA	Modern-Era Retrospective analysis for Research and Applications

Chapter 1

Introduction

1.1 The Threat of Air Pollution: Impact on Health

Air pollution is a complex mixture of particles of varying origin and chemical compositions, with different sources contributing to different sizes of particles¹. Air pollution is the world's largest health threat and is a matter of grave concern for different stakeholders such as policy makers, research groups and people around the world due to the adverse health effects. Annually, 4.6 million life-loss is attributed to air pollution with more than 66% occurs only in Asia² and 1.42-1.92 million deaths are from India³. Air pollutants such as fine particles (particulate matter (PM)), trace gases such as Sulphur dioxide (SO_2), nitrogen oxide ($\text{NO}_x \approx \text{NO} + \text{NO}_2$), Ozone (O_3) are linked to several acute and chronic respiratory diseases. Nearly, 59% of deaths in India are can be traced back to ambient particulate matter pollution³ which are primarily from the combustion of fossil fuels and vehicle exhaust, pose significant health risks¹. The death rate increased in last two decades by 115% - 139% due to ambient PM and O_3 pollution³. Although, respiratory tract and cardio vascular systems are major organs affected by the air pollution, other body organs are also susceptible to the oxidative stress due to increased air pollutant exposure there by leading to systemic inflammatory effects⁴.

1.2 Air Pollution Challenges and Climate Change Implications in Mumbai

Mumbai is the fourth most polluted megacity globally⁵, has been grappling with poor air quality and has recorded high levels of pollution, categorized as "very poor and unhealthy" on the Air Quality Index (AQI) scale⁶. However, during the COVID-19 lockdown, there was a noticeable improvement in air quality, with reduced PM and NO_2 and SO_2 levels⁶. The past year 2022, Mumbai witnessed a noticeable deterioration in air quality in Mumbai, India's financial hub, during the peak winter months of November to January in 2022-23⁷. This decline

in air quality was reflected in the increased number of poor and very poor air quality days, with a 135% increase in extreme bad air quality days compared to previous years⁸. Mumbai, known for its coastal location and natural cleansing advantages, experienced unusual wind patterns and calmer wind spells, attributed to a larger meteorological phenomenon linked to climate change. These changes disrupted the city's natural cleansing mechanism, reducing the dispersal rate of pollutants and contributing to the trapping of high-flying dust emissions⁸. Despite the odds, India is determined to attain net-zero emissions by end of 2070 which is considered as an important move to mitigate climate change (COP26).

1.3 COVID Lockdown Impact in Air Quality

The novel corona virus was declared as “Global pandemic” by World Health Organization on 12 March, 2020⁹. Strict lockdown was imposed in India from 15th March, 2020 to 14th April, 2020 for about 21 days (Table 1). This include travel restrictions for industries, ecommerce, local transport, and aviation, except limited services for milk and groceries and medical emergencies were permitted. Although, this scenario posted a huge threat to India’s economy and well-being, it allows researchers to study about the air quality without pollution emission.

Table 1: COVID-19 Lockdown in India.

Date	Significance
December 31, 2019	China alerts the World Health Organization about the spread of “Viral pneumonia” in Wuhan ¹⁰ .
January 15, 2020	Japan confirms it is the first case of COVID infection ¹⁰ .
January 30, 2020	India confirms its first case of COVID in Thrissur, Kerala ¹⁰ .
March 22, 2020	Janata Curfew: A 14-hour voluntary lockdown was observed in India ^{10,11} .
March 25 – April 14, 2020	Phase 1: National wide lockdown in India Suspension of railways and all flights operations and local transport systems. Local transport was only permitted for essential and emergency services ¹⁰⁻¹² .
April 14 – May 3, 2020	Phase 2: Lockdown extended with categorization into different zones (red, orange, green). Interstate transport allowed for uninfected individuals and small shops were allowed to function with 50% of workers ^{10,11} .
May 4 – May 31, 2020	Phase 3: Lockdown extended with relaxed restrictions. Bus and metro were only limited to certain working groups and time. Weddings and funerals with limited gatherings. Work and education continue to be online with staggered openings ¹⁰⁻¹² .
June 1 – June 30, 2020	Unlock 1: Only Large gatherings were restricted, and night curfews were in effect.
July 1 – July 31, 2020	Unlock 2: Night curfews and limited international travels

August 1 – August 31, 2020	Unlock 3: Relaxed night curfews and all intrastate travel was permitted. However, Maharashtra Government continued to hold the lockdown for the entire month ¹³ .
September 1 – 30, 2020	Unlock 4: Lockdown was imposed only in containment areas.
October 1-31, 2020	Unlock 5: Restaurants, hotels relaxed to operate with 50% capacity. Education institutions remains closed.
April 5 – June 15, 2021	Lockdown 2.0
15 June, 2021	States moved to Unlock phase.

The COVID-19 pandemic and subsequent lockdown measures have had a significant impact on air quality worldwide. Many studies reported improved air quality during a lockdown, as criteria pollutants decreased significantly^{14–17}. Air Quality Index (AQI) for about 21 states and districts was reduced by about 21.28% in Phase1 and further reduced by 10 % in Phases 2 & 3¹⁸. PM and NO₂ remains as most studied pollutants globally and multiple studies assessed the impact of lockdown on these pollutants, since they are majorly emitted by traffic sources and adversely affect the health conditions. 30 -60% and 20 -30% of reductions in NO₂ and PM_{2.5} were reported throughout India. Apart from these pollutants, O₃ contradictorily evidenced increase in concentration in multiple studies¹⁹. While the majority of studies reported reductions in NO₂ and PM_{2.5} concentrations, Gopikrishnan et al. (2022) and Shehzad et al. (2021) observed an increase in O₃ levels during the lockdown. This divergence could be attributed to complex atmospheric chemistry and the interplay between precursor gases and solar radiation. Diminished human activities, including reduced traffic volume and industrial productivity, results in a significant reduction in emissions of NO_x and volatile organic compounds (VOCs), which are precursors to O₃ formation. A study conducted in Changzhou, China, analyzed the characteristics of O₃ precursors during the pre-lockdown, full-lockdown,

and partial-lockdown periods. It was found that the emissions of NO_x and VOCs were substantially suppressed during the full-lockdown period, leading to a significant drop in the concentrations of most VOCs, particularly aromatics which shifted the O₃ formation regime from being VOC-sensitive to a junction of VOC- and NO_x-limited²⁰.

In the context of the MMR region, there is a notable research gap concerning the district-level assessment of air quality during and after the COVID-19 lockdown. This information is particularly valuable for municipal corporations seeking to develop effective strategies. Given the scarcity of ground-based measurements prior to 2020, our study employs satellite data to analyze the post-lockdown impact on air quality trends. Furthermore, we aim to account for the influence of meteorological factors, enabling us to disentangle the effects of air quality changes from those directly attributable to the lockdown measures.

2.1 Objectives of the Report

- Thorough quality examination of Continuous Ambient Air Quality Monitoring Station (CAAQMS) data managed by CPCB/Non-CPCB organisations (SPCBs, IITM, IMD) to ensure proper evaluation of lockdown impact on air quality.
- To evaluate the impact of lockdown in MMR using satellite and ground-based measurement.
- To assess the role of meteorology to modulate the lockdown impact on air quality in MMR.

Chapter 2

Methodology

2.1 Study Area

The Mumbai Metropolitan Region (MMR) encompasses a vast area of 6,328 square kilometers (<https://MMR.maharashtra.gov.in/about-mmr>) and includes nine municipal corporations, namely Greater Mumbai, Navi Mumbai, Kalyan, Ulhasnagar, Vasai, Mira-Bhayandar, Panvel, Bhiwandi-Nizampur, and nine municipal councils, such as Kulgaon-Badalapur, Ambarnath, Matheran, Khopoli, Pen, Alibaug, Karjat, Uran, and Palghar (Figure 2.1). Additionally, it encompasses over 1,000 villages located in the Raigad, Thane, and Palghar districts. The region has a significant population of approximately 23.5 million people, as reported in the 2011 Census. Metropolitan areas like Mumbai have experienced rapid growth in recent decades, leading to a substantial influx of people migrating from rural to urban areas worldwide²¹. One notable characteristic of Mumbai is its relatively high vehicle density, with approximately 1,900 cars per square kilometer, whereas the rest of Maharashtra exhibits a density of 123 vehicles per square km (<https://www.news18.com/news/auto/mumbai-has-highest-vehicle-density-in-india-despite-lesser-cars-than-delhi-2528843.html>). The MMR experiences a monsoon-dominated climate, with a significant amount of rainfall occurring from June to September. The region receives heavy precipitation during the monsoon season, which is vital for agriculture, water resources, and overall ecosystem health. The average annual rainfall in Mumbai is around 2,200-2,500 mm, with variations across different parts of the MMR.

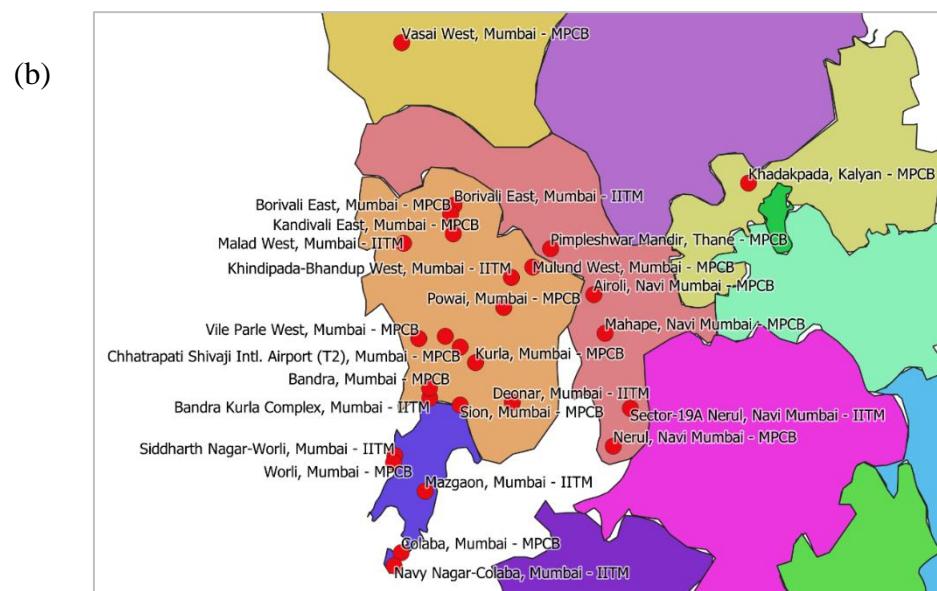
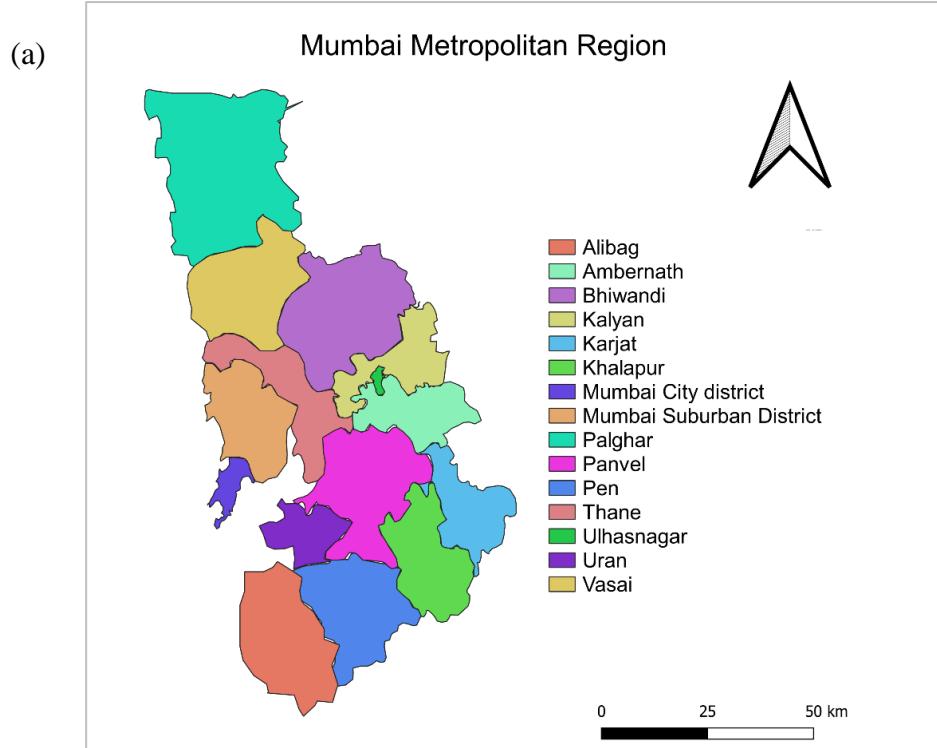


Figure 2.1. CAAQMS stations available in MMR Region. The districts within MMR region are given in different colors in panel (a) and the annotations in the panel (b) denote the CAAQMS station names.

2.2 Mumbai Metropolitan region shape file:

The toposheet for the MMR region (Figure 2.2) was downloaded from Mumbai Metropolitan Region Development Authority website (<https://MMR.maharashtra.gov.in/home>). Georeferencing of the toposheet was performed using QGIS (Fig 2.2a), and the MMR boundary was digitized manually on top of the toposheet (Fig 2.2b).

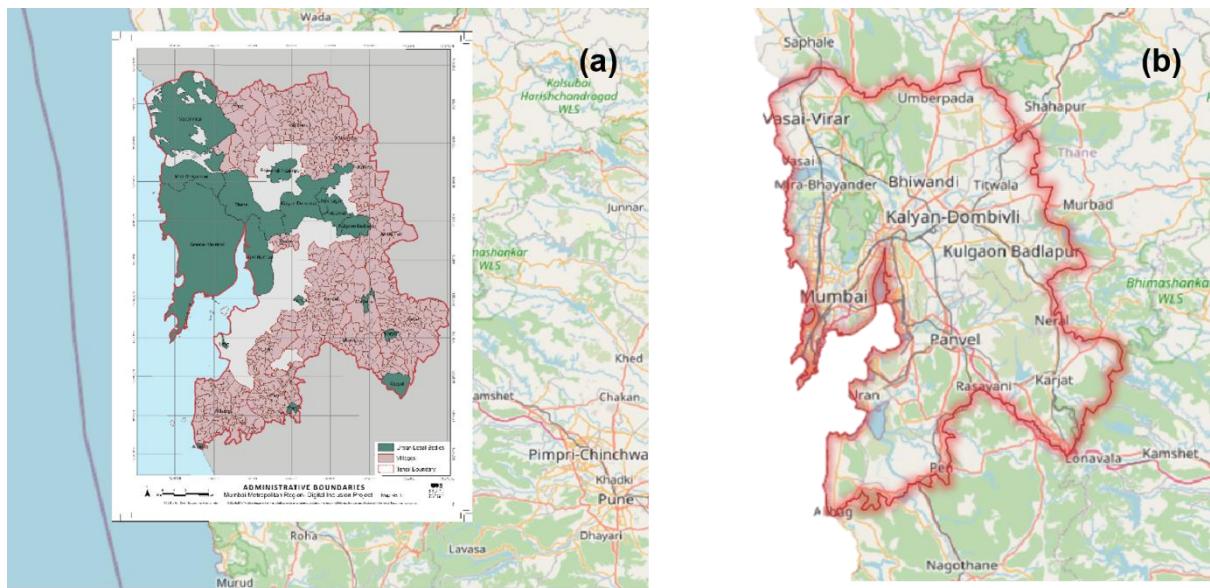


Figure 2.2. Georeferencing of MMR. The toposheet is projected on the OSM standard map with EPSG: 32643 -WGS 84 projection (b). Post georeferencing, the polygon was digitalized over it and exported as shapefile (.shp)

2.3. Air Quality Data

2.3.1 Ground station measurements:

This study utilized the air quality data retrieved from CAAQMS monitoring stations which are available on the central control room for air quality management website (<https://app.cpcbCCR.com/CCR/>). The analysis is performed with data of critical criteria pollutants, such as PM_{2.5}, PM₁₀, NO₂, NO, NO_x, NO₂, and O₃, in 15 minutes for about 27 sites (Fig 2.1) for the years 2020 and 2019 (Table 2.1). The used Python scripts for data cleaning and analysis are made available on GitHub (Chapter 3). Ground station spatial and temporal

data availability is scarce (Table 2.1) and changes with regions and stations, making it difficult to delineate the geophysical long-term trends assessments. We used satellite data in our study to fill the data gap.

2.3.2 Satellite measurements

Various studies used the vertical distribution of trace gases and aerosol information to estimate ground-level pollutant concentrations utilizing chemical transport models. Applying these techniques to complement the ground station measurements would help bridge the data gap in CPCB data. This would, in turn, help derive a meaningful picture of the changes in ground-level concentrations in sites with no ground monitoring. For our initial preliminary assessments, we utilized the surface pollutant estimates that are freely available.

Surface PM_{2.5} Estimates: For our preliminary exploratory studies, we utilized the annual and monthly ground-level particulate matter estimates developed by ²³, which were estimated by combining the Aerosol Optical Depth (AOD) retrievals from MISR, NASA MODIS and SeaWiFS instruments with GEOS-Chem chemical transport model. They also calibrated the model output with ground-based PM_{2.5} measurements. The datasets (<https://sites.wustl.edu/acag/datasets/surface-pm2-5/>) provided by this research group are freely accessible and are provided in NetCDF [.nc] file format in WGS84 projection. Daily AOD retrievals are combined to estimate the best monthly AOD at 0.1° × 0.1° (11.1 km × 11.1 m) using AERONET weighting. Surface PM_{2.5} estimates are based on monthly AOD to monthly PM_{2.5} mean simulated by the GEOS Chem model (<http://geos-chem.org>; V11–01).

$$PM_{2.5} = \eta(x, y, t) \times AOD_{BE}$$

Where, η is spatio-temporally varying ratio between AOD_{BE} and PM_{2.5}.

Table 2.1. CPCB Monitoring sites and Parameters being reported. Pre-COVID studies were only possible for two stations such as Bandra Mumbai MPCB and Pimpleshwar Mandir, Thane – MPCB, since air quality data for 2019 is completely available only for these sites. Source: CPCB

Place	Station	Status	Data availability			
			2021	2020	2019	2018
Kalyan	Khadakpada, Kalyan - MPCB	Live	Jan	All	Jun - Dec	
Mumbai	Bandra Kurla Complex, Mumbai - IITM	Delay	Jan	Nov		
Mumbai	Bandra, Mumbai - MPCB	Inactive	All	All	All	All
Mumbai	Borivali East, Mumbai - IITM	Live	Jan	Nov		
Mumbai	Borivali East, Mumbai - MPCB	Live	Jan	Jan - Apr	Jun - Dec	
Mumbai	Chakala-Andheri East, Mumbai - IITM	Live	Jan	Dec		
Mumbai	Chhatrapati Shivaji Intl. Airport (T2), Mumbai - MPCB	Delay	Jan	Jan - Dec	Jun - Dec	
Mumbai	Colaba, Mumbai - MPCB	Live	Jan	Jan - Dec	Jun - Dec	
Mumbai	Deonar, Mumbai - IITM	Live	Jan	Nov - Dec		
Mumbai	Kandivali East, Mumbai - MPCB	Delay	Jan	Dec		
Mumbai	Khindipada-Bhandup West, Mumbai - IITM	Live	Jan	Dec		
Mumbai	Kurla, Mumbai - MPCB	Live	Jan	All	Jun - Dec	
Mumbai	Malad West, Mumbai - IITM	Live	Jan	Nov - Dec		
Mumbai	Mazgaon, Mumbai - IITM	Live	Jan	Nov - Dec		
Mumbai	Mulund West, Mumbai - MPCB	Live	Jan	Dec		
Mumbai	Navy Nagar-Colaba, Mumbai - IITM	Live	Jan	Nov - Dec		
Mumbai	Powai, Mumbai - MPCB	Live	Jan	All	Jun - Dec	
Mumbai	Siddharth Nagar-Worli, Mumbai - IITM	Live	Jan			
Mumbai	Sion, Mumbai - MPCB	Live	Jan	All	Jun - Dec	
Mumbai	Vasai West, Mumbai - MPCB	Live	Jan	All	Jun, Sept - Dec	
Mumbai	Vile Parle West, Mumbai - MPCB	Live	Jan	All	Jun - Dec	
Mumbai	Worli, Mumbai - MPCB	Live	Jan	All	Jun - Dec	
Navi Mumbai	Airoli, Navi Mumbai - MPCB	Inactive		Jan - March	Jan - Oct	
Navi Mumbai	Mahape, Navi Mumbai - MPCB	Delay	Jan	All	Jun - Dec	
Navi Mumbai	Nerul, Navi Mumbai - MPCB	Delay	Jan	Jan - April, Sept - Dec	Jun - Dec	
Navi Mumbai	Sector-19A Nerul, Navi Mumbai - IITM	Live	Jan	Nov - Dec		
Thane	Pimpleshwar Mandir, Thane - MPCB	Live	Jan	All	All	

TROPOMI Derived NO₂ estimates:²⁴ produced freely available NO₂ surface level estimates using tropospheric NO₂ columns from O₃ Monitoring Instrument (OMI) on NASA's Earth Observing System Aura satellite and Tropospheric Monitoring Instrument (TROPOMI). The TROPOMI has provided NO₂ information since 2018, with finer spatial resolution. Cooper et al. (2020) derived the satellite-derived NO₂ ground-level estimates using the GEOS-Chem chemical transport model.

2.3.3 Meteorological data

ERA5 reanalysis data is available from 1950²⁵. The reanalysis combines the data from the model with the global observations. It estimates many atmospheric parameters, such as ocean waves, wind speed, and velocity. The 10m u-component of wind and 10m v-component of wind were retrieved from ERA5 Hourly data Reanalysis with a spatial resolution of 0.25° × 0.25° at an hourly level and utilized for studying the wind pattern over the MMR region.

2.4 Methodology and Analysis

The bulk data from the CPCB server was automated to download the real-time data using the web scraping method with the help of the Selenium package in Python 3.3. (<https://docs.python.org/3.3/>). However, when the server is unresponsive, manually the data is retrieved. For all 26 sites in MMR regions in Mumbai, Navi Mumbai, Thane, Airoli, and Khadakpada, 15 mins of raw data were collected for 2019, 2020 and 2021 for all criteria pollutants available in the station.

The surface PM_{2.5} and NO₂ estimates from satellite data are retrieved from studies by Van Donkelaar et al. (2021) and Cooper et al. (2022). The data points within the MMR region boundary were utilized for further calculations and visualizations.

2.4.1 Methodology for Comparing Satellite data with CPCB Data

PM_{2.5} and NO₂ for April 2020 and 2019 were considered for the comparative studies. The satellite estimates were only available at the monthly level for 2020 and 2019. Only April 2020 was in complete lockdown in India, and henceforth all our comparison studies utilized the data from April 2020. The CPCB data was initially cleaned using three standard deviations, and the daily mean calculated, from which the monthly mean was calculated. All the missing values were removed from the studies. Data gaps of less than 30 mins were interpolated using the local mean. Data gaps for more than 30 mins were removed from the analysis.

In the satellite products, 16 grid cells around each CPCB station's location were collected and exported for April. The data were compared by comparing the monthly variation throughout the year with the values in 16 grid cells. For comparison of PM_{2.5} values, the CPCB station's 24-hour mean values were used. In the case of NO₂, only the site's value within the satellite overpass time (1:30 pm IST) was utilized.

2.4.2 Methodology for estimating wind speed from reanalysis data:

The hourly 10m u-component of wind and 10m v-component of wind were retrieved with a spatial resolution of 0.25° × 0.25°. Using the u and v components, the wind speed and direction were calculated at the hourly level. The hourly level data are averaged to get daily level averages and monthly level averages ²⁶.

$$\text{Wind Speed} = \sqrt{u^2 + v^2}$$

$$\phi = \text{mod} \left(180 + \frac{180}{\pi} \text{atan2}(u, v), 360 \right)$$

2.4.3 Statistical Analysis

Ordinary least square regression against satellite estimates and CPCB values was performed to compare the values. After preliminary visualizations of the satellite products at the monthly level, all the data points within the MMR boundary were paired against each other for 2020 and 2019 year for March, April, and May months. A paired-T test is used to confirm the response of lockdown in the pollutant data.

We utilized the O₃, SO₂, and NO₂ total column data from OMI, Merra-2, and TROPOMI to visualize the long-term trends of the pollutants from 2017 – 2022. The raw retrievals were visualized and compared for April of the year.

Ordinary least square regression against satellite estimates and CPCB values was performed to compare the values. After preliminary visualizations of the satellite products at the monthly level, all the data points within the MMR boundary were paired against each other for 2020 and 2019 year for March, April, and May months. A paired-T test is used to confirm the response of lockdown in the pollutant data.

We utilized the O₃, SO₂, and NO₂ total column data from OMI, Merra-2, and TROPOMI to visualize the long-term trends of the pollutants from 2017 – 2022. The raw retrievals were visualized and compared for April of the year.

2.4.4 Business as Usual analysis

We cleaned the PM_{2.5} from CAAQMS stations (PM_{2.5}, GM) in MMR region for 2019-2021 years and regressed against the PM_{2.5, sat} at district level. We improvised the concentration dependent error bias in PM_{2.5, sat} by weighting with PM_{2.5, GM} to get PM_{2.5, adj sat}. We used the PM_{2.5, adj sat} to train a random forest model between 1994-2018 years and tested the model prediction efficiency on 2019 monthly data. Post parameter tuning, we used the developed model to predict Business as Usual (BAU) for 2020 and 2021 COVID years (Figure 2.3).

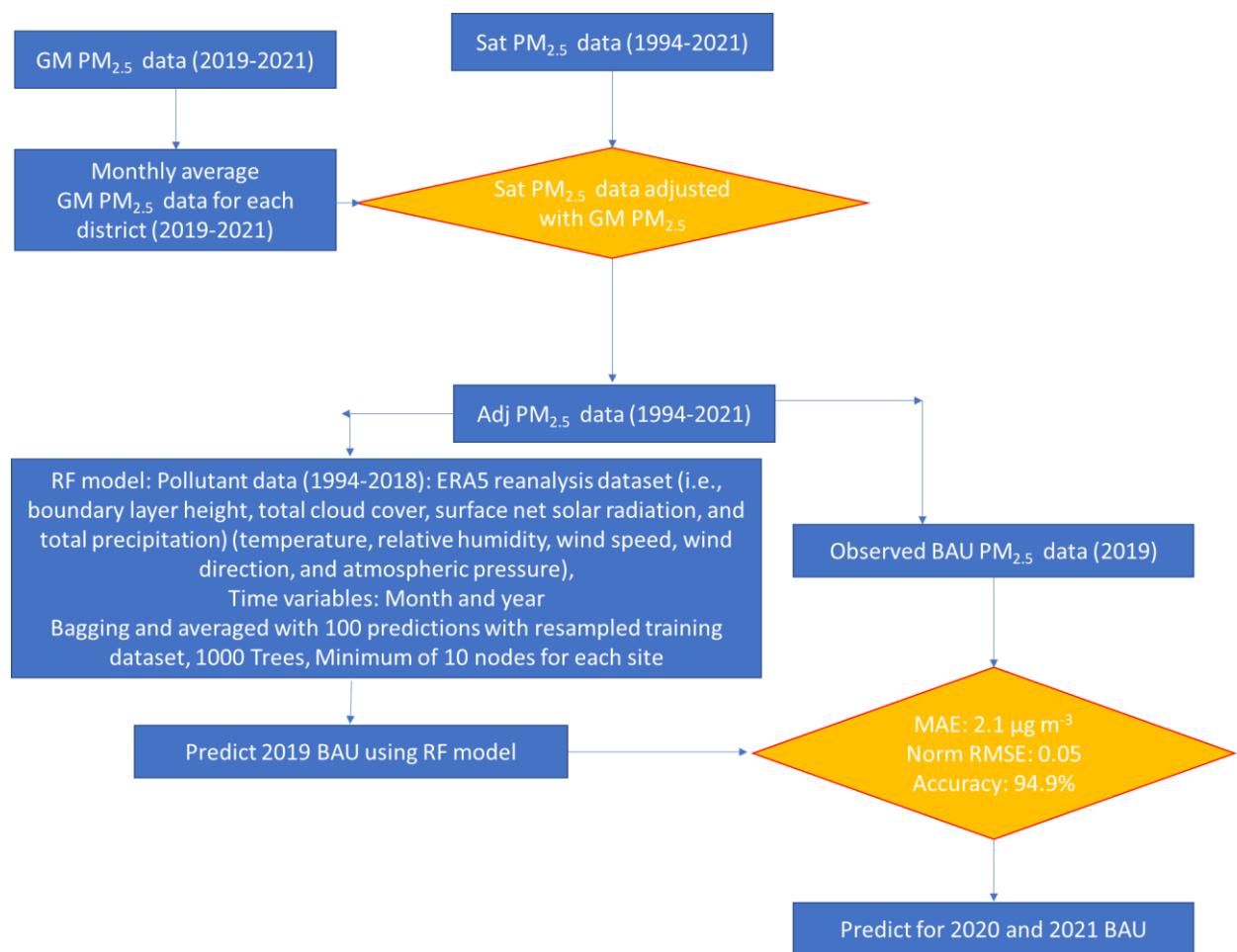


Figure 2.3. Methodology for Business-as-Usual prediction for 2020 and 2021.

We used Theil-Sen, a non-parametric method used for estimating the slope of a linear relationship and to identify the significance of trends in PM_{2.5} values in BAU and actual observed values. It is a robust alternative to ordinary least squares (OLS) regression, which is

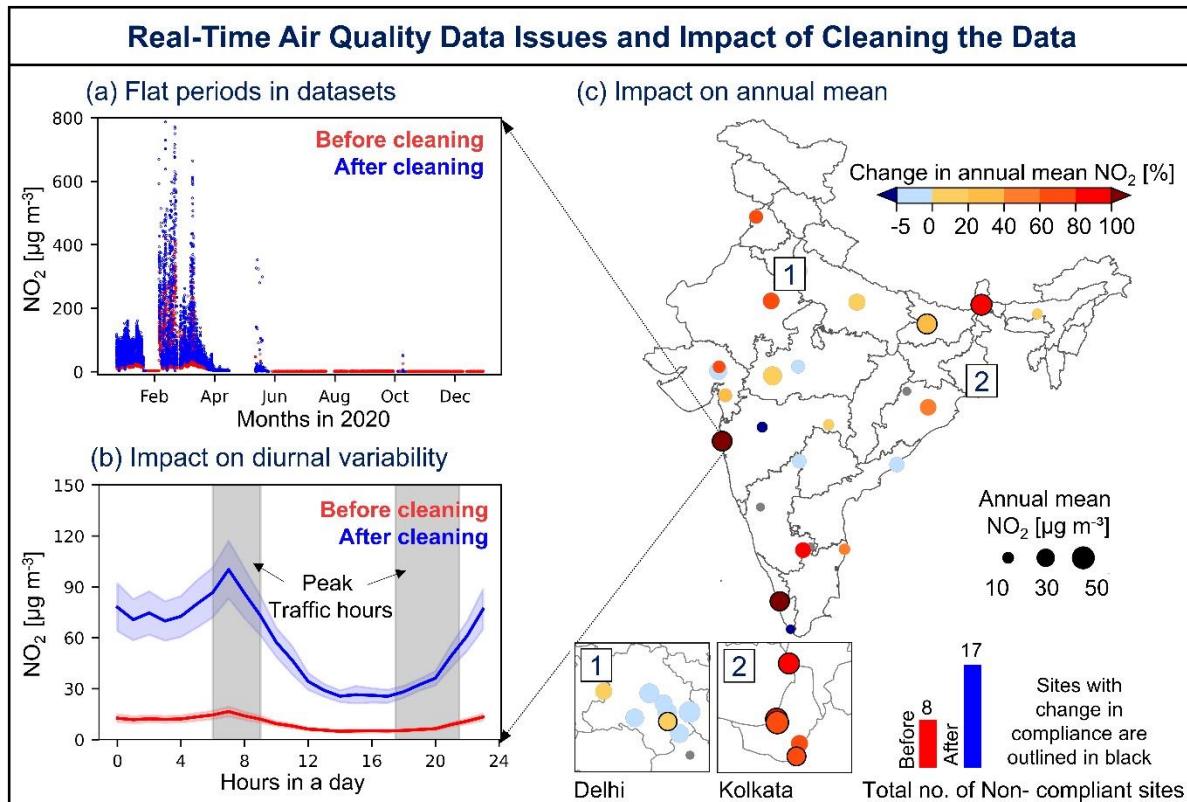
sensitive to outliers in the data. The Theil-Sen estimator provides a robust estimate of the slope by calculating the median of all possible pairwise slopes between data points (<https://search.r-project.org/CRAN/refmans/robslopes/html/TheilSen.html>).

We use Conditional Probability Function (CPF) a statistical function from OpenAir package (<https://davidcarslaw.com/files/openairmanual.pdf>) in R programming to understand the impact on wind direction which is associated with high concentration of PM_{2.5}. It is performed by counting the number of instances where each wind direction bin (e.g., North, Northeast, East, etc.) is associated with a high concentration of the pollutant and probability is calculated by dividing the count for each bin by the total count of high pollutant concentration events.

We use feature importance study to score different input features for a BAU model. A higher score in feature importance analysis means that specific feature is dominant in predicting a certain predictor variable, the model is built for. We used the inbuilt feature importance (https://inria.github.io/scikit-learn-mooc/python_scripts/dev_features_importance.html) by forest model in sklearn.

Chapter 3

Addressing the CAAQMS Data Quality Issues



3.1 Necessity of assessing the data quality

Air pollution is the single largest environmental risk to human health in India (Pandey et al., 2021b), responsible for 1.5-2.5 million deaths each year^{28,29}. Air quality in India has deteriorated in the last two decades (Dutta & Chatterjee, 2022), rapidly in the fast-growing cities ³¹. The National Clean Air Programme (NCAP) was launched in 2019 to reduce particulate matter (PM) pollution by 20-30% in 122 Indian cities by 2024 and this has led to an extensive real-time air quality monitoring network³² to ensure regulatory compliance with National Ambient Air Quality Standards (NAAQS)³³. The number of Continuous Ambient Air Quality Monitoring Sites (CAAQMS) maintained by the Central and State Pollution Control Boards (Hereafter referred to as CPCB and SPCB, respectively) of India has increased from 30

in 2013 to 432 in 2023³⁴. This has led to the dramatic growth in the application of the CAAQMS datasets globally as evidenced by over 600 studies on Google Scholar (last accessed 10 March 2023) found using the keywords "CPCB", "CAAQMS", "air pollution" and "India". These use the CAAQMS datasets to examine spatiotemporal variability in air pollution^{35–37}, identify sources of pollution^{38,39}, test the effectiveness of pollution control policies and measures^{40,41}, assess satellite data products and atmospheric modelling outputs^{42–44} and estimate the health risk from air pollution^{27,45,46}. Half of the studies using the CAAQMS datasets were published in the last 2-3 years and assessed the impact of COVID-19 lockdown on air quality. CAAQMS data-derived metrics such as the Air Quality Index (AQI) are used by environmental activists to advocate for better policies and engage the public^{47–49}. The credibility of this data is thus of paramount importance to these stakeholders, but there are no official procedures in place for data ratification⁵⁰. This has led to increased data quality concerns, such as continuously repeating observations^{51,52} and inconsistencies in reporting units⁴² and data records^{53–60} limiting the utility of CAAQMS data. Recent studies use novel techniques to remove spurious observations from the CAAQMS data^{39,42,52}, but there are no comprehensive publicly available tools to provide quality-assured datasets of air quality.

In this study, we develop the first automated tool “AirPy” which identifies and addresses a series of data quality issues in the CAAQMS datasets. We further discuss the implications of using our proposed solution on air quality and related metrics, which play a vital role in shaping policy decisions and are beneficial for various stakeholders such as research organizations and institutions, environmental regulatory agencies, judicial bodies as well as media agencies. Our automated tool will streamline the process of data cleaning and benefit this diverse group of end-users of the CAAQMS datasets.

3.2 Methodology of AirPy data cleaning tool

3.2.1. Pollutant data

We use open-source real-time air quality data for CAAQMS from the Central Control Room dashboard (CCR: Central Control Room for Air Quality Management - All India.): <https://app.cpcbCCR.com/CCR/#/caaqm-dashboard-all/caaqm-landing>; last accessed: 31 October 2022). We retrieve data for six air pollutants — fine and coarse particulate matter (PM_{2.5} and PM₁₀), nitric oxide (NO), nitrogen dioxide (NO₂), nitrogen oxides (NO_x), and O₃ (O₃) at 15-minute resolution for the three-year period, 2019-2021. We download data for 40 sites, representing ~25% of all CAAQMS operational in those years (CPCB, 2021) and spanning across the entire country (Figure 3.1). These years provide a unique opportunity to investigate the data quality issues with the sites newly set up as part of NCAP and examine if these issues are linked to the pandemic or have persisted throughout.

3.2.2. Data quality control

We develop a tool for data-driven air quality data cleansing using Python (V3.7; <https://www.python.org/ftp/python/3.7.3>; last accessed: 17 March 2023) along with widely used data science libraries such as Pandas and NumPy. The tool has a sequential process to address issues such as consecutive repeats, outliers, continuous negative values, and unit inconsistencies. “Consecutive repeats” are defined as persistent and inconsistent fluctuations over a duration exceeding 24 hours. To address this, we remove datasets with a coefficient of variation (CoV, ratio of standard deviation to the mean in a 24-hour running window) less than 0.1. We test the sensitivity of this threshold and find 0.1 to be optimal. We evaluate the robustness of this threshold, we select five pollutant records from five sites that contained consecutive repeats. We manually label each 15-minute data record as either a repeat or non-repeat and vary the threshold from 0.01 to 0.25 comparing the actual label with the label from

our algorithm to calculate accuracy. Our results show that a threshold of 0.1 had an optimally high accuracy of 91% (Figure 3.2).

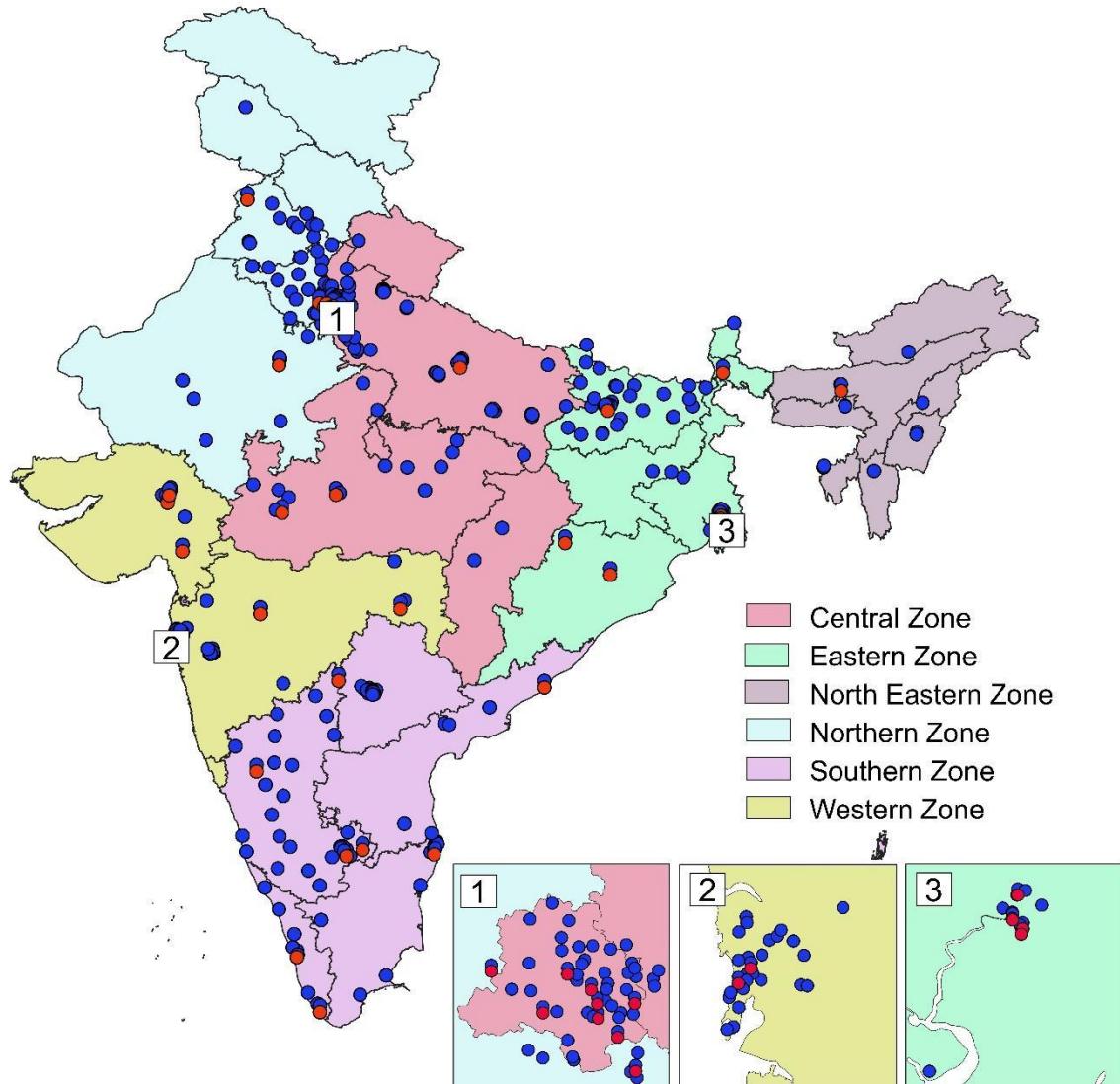


Figure 3.1: Spatial coverage of Continuous Ambient Air Quality Monitoring Stations (CAAQMS) in India. We study 40 sites operational in 2019-2021 (red) and rest of the sites are shown in blue. Insets show the zoomed-in maps for Delhi (1), Mumbai (2), and Kolkata (3). State boundaries are from GitHub repositories^{62,63}.

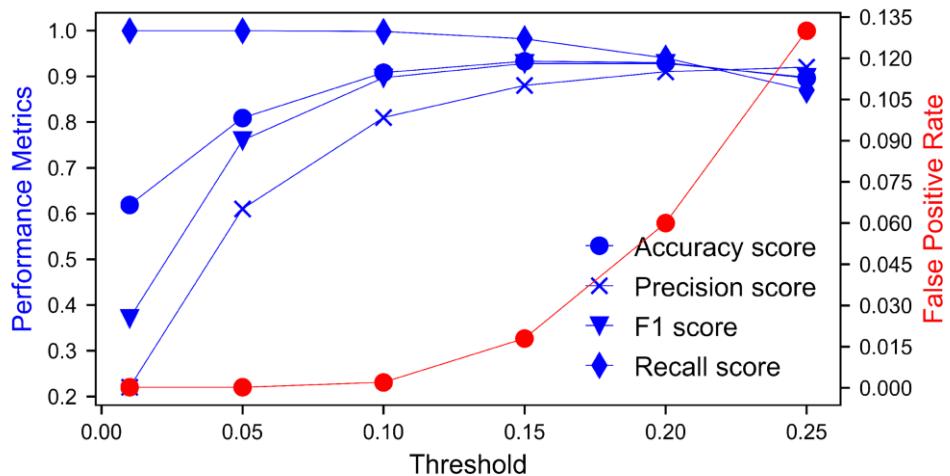


Figure 3.2. Determining the optimal threshold for the consecutive repeat algorithm. The performance metrics, including accuracy score, precision score, F1 score, and recall score, are plotted against the y-axis on the left for different thresholds in blue color, while the false positive rate is plotted against the y-axis on the right for different thresholds in red color.

In comparison, Singh et al. (2020) used a threshold of 0.05⁵² with an accuracy of only 81%, and Silver et al. (2018) employed thresholds manually specific to each dataset⁶⁴ ranging from 0.03 to 0.05 which results in accuracy score between 70% and 81% in our sample set (Figure S15). We improvised this existing methodology by rigorously testing the running window of the timeseries and thresholds which surpasses the approaches employed in previous air quality studies^{64–66} with an exceptional accuracy score. These studies solely remove consecutively repeating absolute values for > 24 hours, which has an accuracy of only 57.4% or > 4 hours which has accuracy of 60.9% within the sample set considered in this study. These findings suggest that a CoV threshold of 0.1 can effectively identify and remove repeated observations in a dataset. Accuracy score increases even beyond the threshold of 0.1 but it also leads to an increase in the valid entries being misclassified as consecutive repeats (Figure S15). So, we restrict it to 0.1 to be optimal and prevent increase in the false positive rate.

$$CV_t = \frac{\text{standard deviation for data in the running window of time } t, (\sigma_t)}{\text{mean for the date in the running window of time } t, (\underline{x}_t)}$$

We remove outliers by following the approach of Iglewicz & Hoaglin (1993) (Iglewicz & Hoaglin, 1993) which flags any observation as an outlier if its absolute difference from the median exceeds three times the median absolute deviation (MAD, median distance between each observation and the median of all observations). Here, we calculate the median and MAD for a 3-hour running window.

$$\text{Median Absolute Deviation, } MAD_d = \text{median}(|x_i - \tilde{x}|)$$

$$\text{Outlier score, } S_i = \frac{x_i - \tilde{x}}{MAD_d}$$

Next, we consider two instances of negative observations as invalid. First, PM less than $-10 \mu\text{g m}^{-3}$ following the guidelines of US EPA (Hanley, 2016). Second, any four continuous negative observations result in negative hourly averages as suggested by the instrument manufacturers (Hart, 2010). These invalid negative observations are removed, and we retain any other instance of negative observations to avoid biasing the data low. We also remove datasets where $\text{PM}_{2.5}$ exceeds PM_{10} that are due to irregular calibrations and are not valid (Yao et al., 2009).

The CPCB parameter reporting protocol (CPCB, 2015a) requires each SPCB to report NO and NO_2 as mass concentrations in $\mu\text{g m}^{-3}$ and NO_x as mixing ratio in ppb while transmitting in the CCR dashboard. We address the discrepancy in reporting units by utilizing the ratio of NO + NO_2 to NO_x , along with empirically determined upper and lower bounds, to categorize the data sets based on their actual units as C1, C2, C3 and C4 (Table 3.1). The upper and lower bounds for these categories are listed in Table 3.1. We discover that the data sets in C1 adhere to the Central Pollution Control Board (CPCB) protocol's unit specifications, while those in C2 have NO and NO_2 measurements in ppb instead of $\mu\text{g m}^{-3}$. Data sets in C3 have NO measurements in ppb and NO_x measurements in $\mu\text{g m}^{-3}$, while data sets in C4 have either

NO or NO_x measurements in ppb or $\mu\text{g m}^{-3}$ that were misreported (Table 3.1). To ensure consistency and accuracy, we adjust the reported NO, NO₂, and NO_x values using unit conversion factors at 25 °C of 1.23 $\mu\text{g m}^{-3}$ per ppb for NO and 1.88 $\mu\text{g m}^{-3}$ per ppb for NO₂. We find that temperature does not have a significant effect on the unit conversion factors (Figure S16). Our study builds upon the unprecedented research conducted by Vohra et al. (2021), who were the first to reveal the presence of the C2 category of unit inconsistency within CAAQMS NO₂ data for a site in Delhi ⁴². This is important in shedding light on an aspect of data quality issue in CAAQMS that had been rarely discussed before. In line with this pioneering work, our study goes beyond by not only confirming the presence of unit inconsistencies in the dataset related to nitrogen oxides but also identifying two additional categories of such inconsistencies in CAAQMS data throughout the country, which have not been previously documented or explored in air quality research. Moreover, we have developed a unique tool that automates the clustering of dubiously reported data points expressed in different units. This tool further extends the novelty of our work by enabling the conversion of these data points into units defined by the CPCB, ensuring adherence to standardized reporting protocols with minimal user intervention.

We compute the AQI on an hourly basis using the guidelines set by CPCB. To calculate the AQI, a minimum of three pollutant concentrations, including at least one of PM_{2.5} or PM₁₀, is required. In our analysis, we use NO_x, PM_{2.5}, and O₃ mass concentrations in micrograms per cubic meter ($\mu\text{g m}^{-3}$). As per the CPCB protocol, NO_x is reported in ppb, so we convert it to $\mu\text{g m}^{-3}$ at standard temperature. We calculate the 24-hour running window mean for NO_x and PM_{2.5}, and the maximum 8-hour running window mean for O₃. This is followed by calculating the sub-index (I) for each pollutant (i) with pollutant concentration (C_p) as per CPCB guidelines ⁷² using Equation S1, where I_{HI} and I_{LO} are the AQI values for the upper and lower breakpoint concentrations, and B_{HI} and B_{LO} are the upper and lower breakpoint concentrations,

respectively. Breakpoint concentrations and AQI corresponding to breakpoint concentrations are constants given in Greater Vancouver AQI¹.

$$I_i = [\{ (I_{HI} - I_{LO}) / (B_{HI} - B_{LO}) \} * (C_p - B_{LO})] + I_{LO} \quad (S1)$$

After calculating the sub-indices, we use Equation S2 to determine the AQI (https://app.cpcbccr.com/CCR_docs/AQI%20-Calculator.xls; last accessed: 18 January 2023), which takes the maximum of the sub-indices for all the pollutant concentrations.

$$AQI = \text{Max } (I_p) \text{ (where; } p = 1, 2, \dots, n; \text{ denotes } n \text{ pollutants)} \quad (S2)$$

Table 3.1. Categories of unit inconsistency identified across all sites during 2019, 2020 and 2021. The central tendency for $(NO_2 + NO)/NO_x$ ratio for a particular category is given along with the superscript and subscript which indicates the empirically determined upper and lower limit to categorize the data into four categories.

Category	Central tendency with upper and lower bounds	Actual reporting units		
		NO	NO ₂	NO _x
C1	1.75 ^{1.85} _{1.25}	µg m ⁻³	µg m ⁻³	ppb
C2	1.00 ^{1.25} _{0.8}	ppb	ppb	ppb
C3	0.74 ^{0.8} _{0.5}	ppb	µg m ⁻³	µg m ⁻³
C4	1.90 ^{2.00} _{1.85}	µg m ⁻³	µg m ⁻³	µg m ⁻³

Table 3.2. Unit inconsistency for oxides of nitrogen across sites in India during 2019, 2020, and 2021. The table displays sites reporting oxides of nitrogen in accordance with the CPCB protocol (green) and those that do not comply (C2 in red). Sites dubiously reporting in more than one unit during the year are in yellow. NA refers to no data for the year.

Station name	Zones	2019	2020	2021
CI1	Central India	✓	✓	✓
CI2		✓	✓	✓
CI3		✓	✓	✓
CI4		✓	✓	✓
EI1	Eastern India	✓	✓	✓
EI2		✓	✓	✓
EI3		✓	✓	✓
EI4		✓	✓	✓
EI5		✓	✓	✓
EI6		✓	✓	✓
EI7		✓	✓	✓
EI8		✓	✓	✓
EI9		✓	✓	✓

Station name	Zones	2019	2020	2021
NEI1	North eastern India	✓	✓	✓
NI1		✓	✓	✓
NI2		✓	✓	✓
NI4		✓	✓	✓
NI5		✓	✓	✓
NI6		✓	✓	✓
NI7		✓	✓	✓
NI8		✓	✓	✓
NI9		✓	✓	✓
NI10		✓	✓	✓
NI11		✓	✓	✓
NI12		✓	✓	✓
NI13		✓	✓	✓
NI14		✓	✓	✓
NI15		✓	✓	✓
SI1	Southern India	✓	✓	✓
SI2		✓	✓	✓
SI3		✓	✓	✓
SI4		✓	✓	✓
SI5		NA	✓	✓
SI6		NA	✓	✓
SI7		✓	✓	✓
WI1	Western India	✓	✓	✓
WI2		✓	✓	✓
WI3		✓	✓	✓
WI4		✓	✓	✓
WI5		✓	✓	✓

3.3 Data quality issues and implications of cleaning

3.3.1 Uncovering the persistent data quality issues in CAAQMS data

In our analysis, we consider sites with more than 75% of data available for all atmospheric constituents. We study the implications of our data cleaning tool on air quality metrics such as annual mean, compliance to NAAQS, NO_x-to-NO₂ ratio, and AQI across all three years. Figure 3.3 illustrates instances of three major data quality issues identified across 2019-2021. These are consecutive repeats, outliers, and inconsistent reporting units. On average, we observed 9.2% of trace gases and 2.1% of PM datasets as consecutive repeats (Figure 3.2a).

The frequency of consecutive repeats increased significantly in 2021 by 74% (p-value <0.05) when compared to 2019. Sharp increases (95%) were seen after the initial COVID-19 lockdown in April 2020 with respect to March (Figure 3.4), unlike in 2019 and 2021. We observe a significant negative correlation (r -value =-0.63, p-value <0.01) between the percentage of consecutive repeats and community mobility data (Mathieu et al., 2020) (Figure 3.5) for 2020, suggesting that lockdown restrictions could have influenced data quality. We still see an increase in percentage of consecutive repeats even when the restrictions were reduced in 2021 (Figure 3.4g). We also observe a peak in the frequency of consecutive repeats in July-September each year (Figure 3.4g). These consecutive repeats can be attributed to improper and infrequent calibrations and instrument malfunctions such as photomultiplier tube degradation, changes in calibration (Bhawan & Nagar, 2011), and variations in airflow (CPCB, 2018). We observe high percentages of consecutive repeats at some sites (Figure 3.2a). These range from 5-60% for PM records at 5 sites and to 10-88% for trace gases at 19 sites. Values such as 5 $\mu\text{g m}^{-3}$ and 6 $\mu\text{g m}^{-3}$ are commonly observed repeating consecutively and account for 9.2% of annual data at two sites in southern and western India. We observe recurring fluctuations in O₃ at one site in western India. These are between 34 $\mu\text{g m}^{-3}$ and 62 $\mu\text{g m}^{-3}$ and

have a standard deviation of $3.3 \mu\text{g m}^{-3}$. Our tool successfully removes these fluctuations which are 29.6% of the O_3 data in 2020 (Figure 3.6).

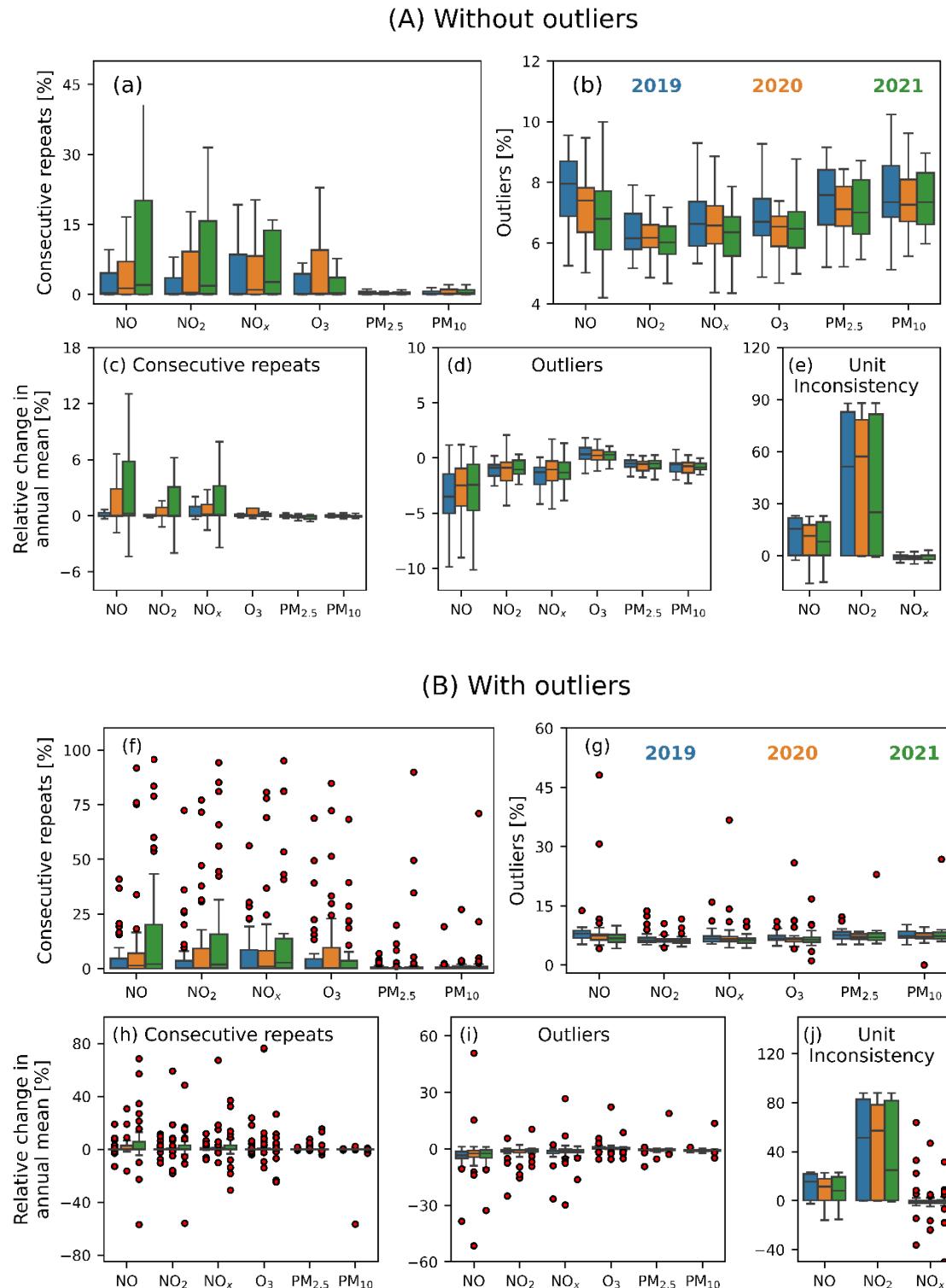


Figure 3.2. Percentage of invalid data and its impact on annual mean of pollutant levels for 2019-2021. Panel B (f-j) shows the same data in Panel A (a-e) but with outliers. Panel (a-b)

show the percentage of consecutive repeats (a) and outliers (b) removed from the total available data for each site in 2019-2021. Here, consecutive repeats and outliers are two exclusive events. Panels (c-d) demonstrate the relative change in annual mean after the removal of consecutive repeats (c), outliers (b), and after ensuring unit consistency (d). The data pertaining to different years are distinguished by bar colours, 2019 in blue, 2020 in orange, and 2021 in green, and each outlier data point in red represents a specific site.

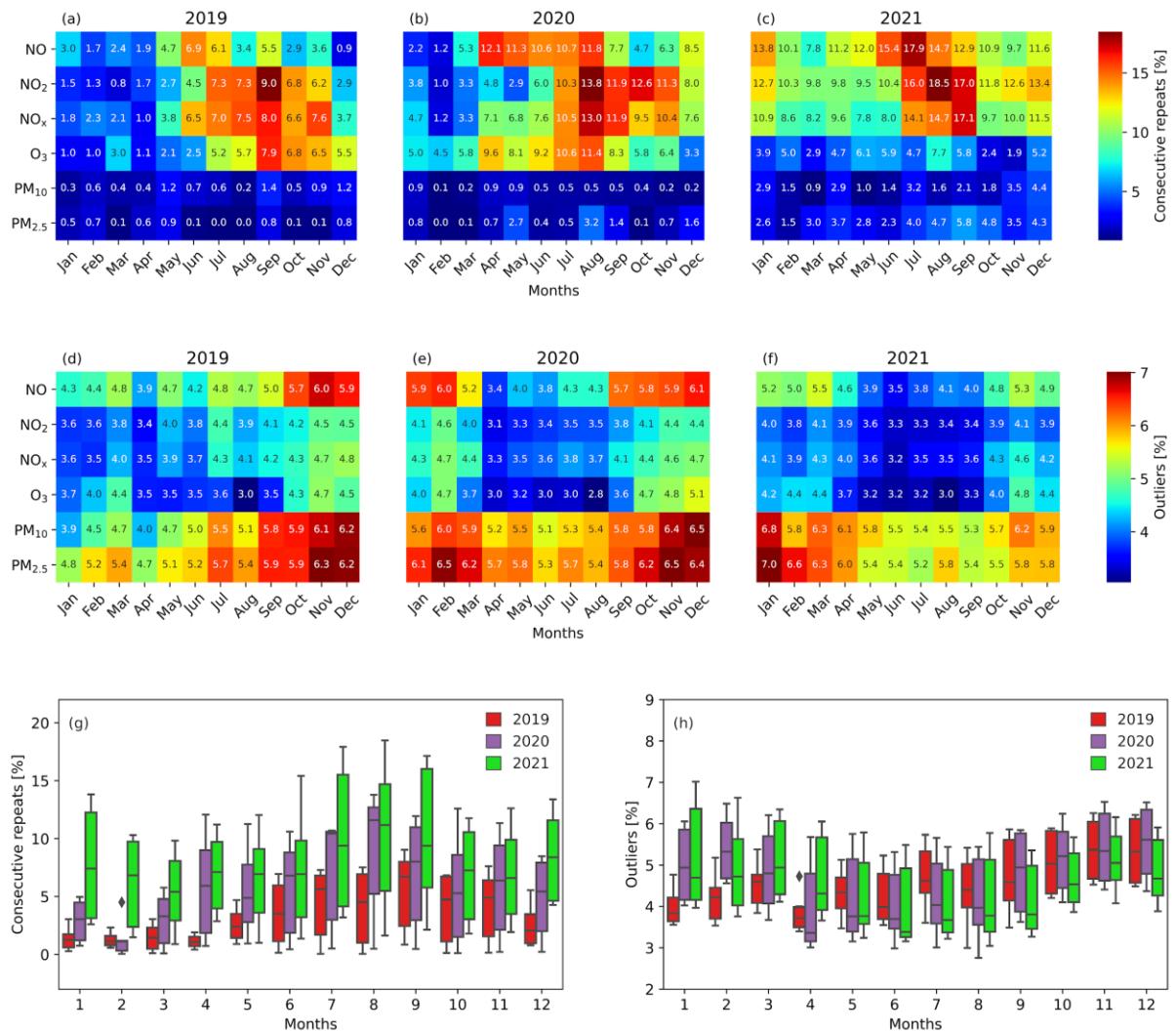


Figure 3.4. Monthly trends in the percentage of data flagged as consecutive repeats and outliers across 2019-2021 for all six pollutants. The percentages of data categorised as consecutive repeats (panels a-c) and outliers (panels d-f) are calculated by considering the total available data as the denominator and normalised across months and sites. The box plots in panels (g) and (h) show the distribution of these percentages for each year across the twelve months.

and (h) illustrate the monthly trends in the percentage of consecutive repeats and outliers for all three years, with the box color indicating the year. Each data point in the box plot represents the normalised percentages for each site in each month across all three years.

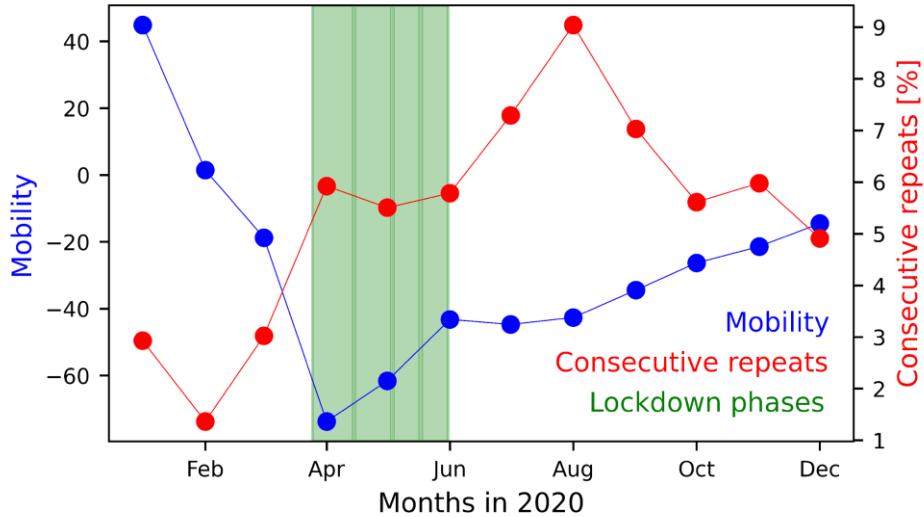


Figure 3.5. Correlation between Mobility and Consecutive Repeats of Air Quality Data during Lockdown Phases in India. The plot presents a correlation analysis between mobility and the consecutive repeats of air quality data during the 2020 lockdown phases in India. The graph shows the average mobility in blue on the x-axis and the normalized percentages of consecutive repeats of air quality data in red on the y-axis and the different lockdown phases (phase 1 to phase 4) are marked by green planes.

We observe outliers in pollutant records (Figure 3.2b) accounting for an average of 7.3% of the datasets across all three years and all monitoring locations. Six sites are identified with outliers exceeding 10% in at least one pollutant record during a particular year. Our analysis shows that most outliers occur between November and March (Figure 3.4h). Our findings also reveal the presence of spurious outliers that are either very low ($<1 \mu\text{g m}^{-3}$, mostly in trace gases and accounting for 0.7% of the data), or very high ($>500 \mu\text{g m}^{-3}$, mostly in PM and for 0.4% of the data). Extreme outliers are the observations which are more than $1.5 \times \text{IQR}$ (interquartile range) of the outliers⁷⁶. The latter case often exhibits readings such as 999.99,

995, and $1000 \mu\text{g m}^{-3}$ (Figure 3.7a) and these do not happen around high pollution events such as dust storms or festival celebrations. At several sites, these outliers recur at the same time interval.

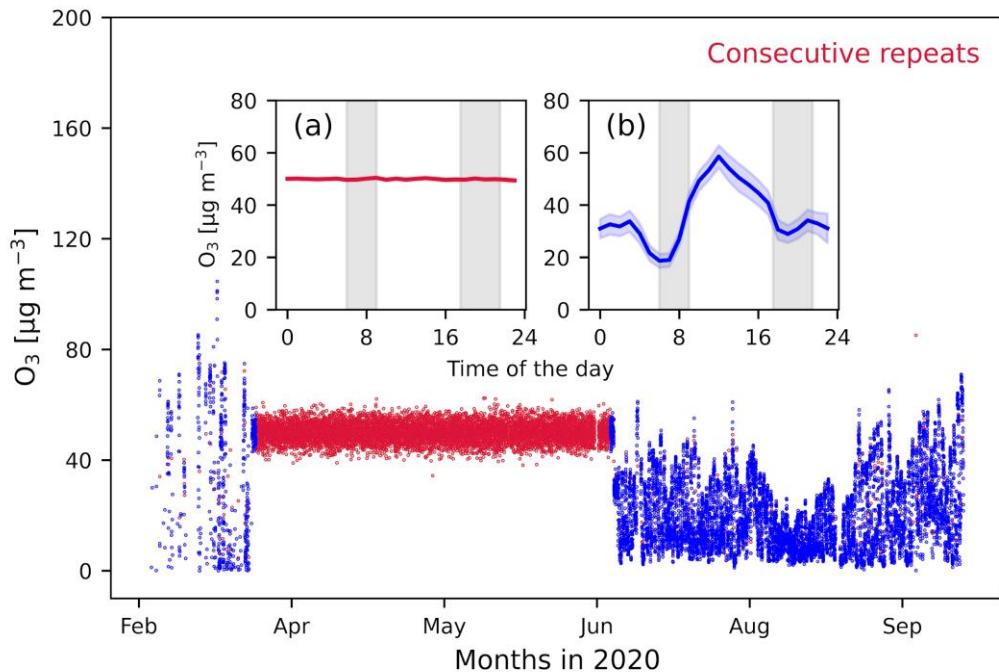


Figure 3.6. Identification of data with no diurnal variability using our algorithm. The primary panel presents a time series of a 15-minute O_3 dataset from a monitoring site in western India. The main panels highlight the repetitive fluctuations removed by the tool in red and its corresponding diurnal variability in inset (a). The other data are shown in blue in main panel with the valid diurnal variation in inset (b).

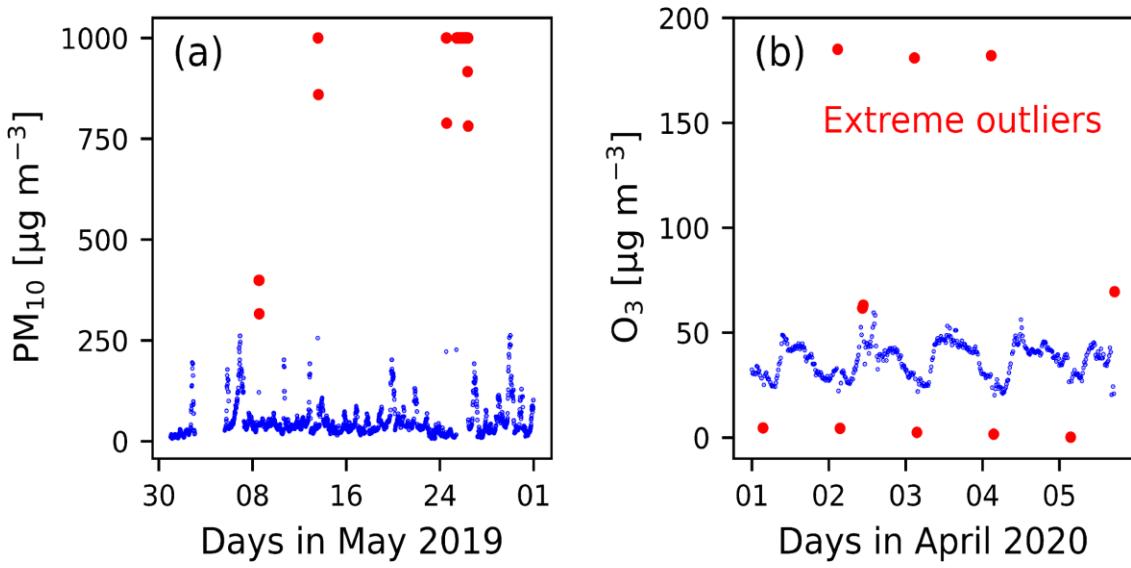


Figure 3.7: Identification of Extreme Outliers in Pollutant Data. Each data point corresponds to the 15-minute average, with extreme outliers marked in red.

Spurious NO₂ observations were recorded consistently at 12:00 IST every day for a site in southern India (Figure 3.3). We suspect that the periodicity in outliers could indicate self-radio frequency interference within the measuring device (Cybersecurity and Infrastructure Security Agency, 2020). Instrument manufacturers (Hart, 2013) suggest that drift in instrument temperature could have caused these anomalies. Our tool classifies these invalid observations as outliers and removes these in accordance with the US EPA (2018) (United States Environmental Protection Agency, 2018).

Out of the 40 sites, 28 reported PM_{2.5} concentrations exceeding PM₁₀, with a maximum difference in the range of 691–921 µg m⁻³, across the three years. On average, this difference increased from 20.4 µg m⁻³ in 2019 to 57.6 µg m⁻³ in 2020 and statistically significant (*p*-value <0.05) increased to 58.9 µg m⁻³ in 2021. Five sites persistently reported PM_{2.5} levels exceeding those of PM₁₀ by more than 100 µg m⁻³ in all three years, while 20 sites reported such differences in at least one study year. These observations were commonly reported at 23:30 IST, occurring 15 times more frequently than at other times of the day. Furthermore, negative

observations (<0.01% of data) in PM_{2.5} and NO at two monitoring sites in northern India due to improper instrument calibration (Hart, 2010) were eliminated.

Figure 4.1c shows a site in north India where NO and NO₂ levels are recorded in $\mu\text{g m}^{-3}$ adhering to the CPCB parameter reporting protocol (CPCB, 2015a) before August 21, 2020, but in ppb after that. Only 11 out of 40 sites consistently adhered to the protocol and 12 sites misreported both NO and NO₂ levels in ppb continuously during 2019-2021 (Table 3.2). Additionally, we detected intricate discrepancies with four different unit combinations in 17 out of 40 sites that reported nitrogen oxides in non-uniform unit combinations (Table 3.2 and Figure 3.8). These could potentially result in the misinterpretation of comparative analyses conducted for the same site and so we rectify this inconsistency in units (Figure 3.9) using our tool.

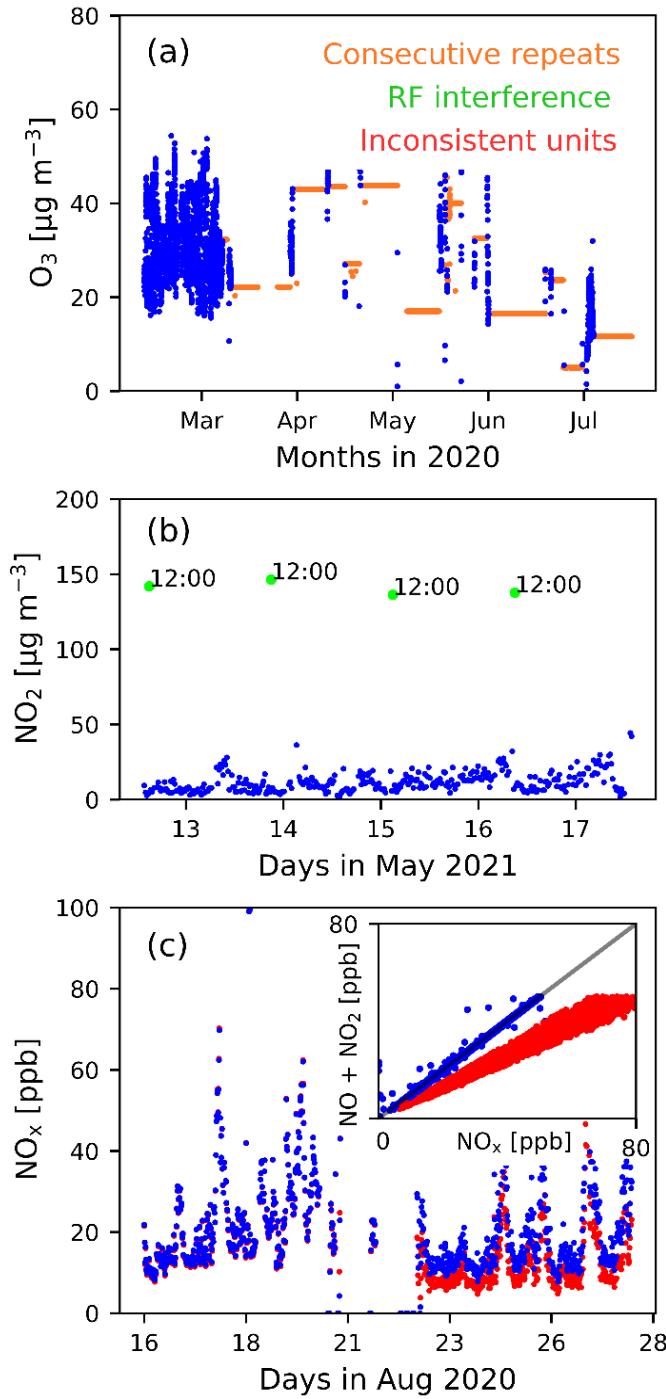


Figure 3.3. Major data quality concerns in CAAQMS data in India. All panels show a time series of 15-minute records. Raw data in all panels are shown in blue. Consecutive repeats are in orange in panel (a) and outliers occurring at the same time each day are in green in panel (b). Panel (c) shows the inconsistency in reporting units. Data records after 21st August 2020 have a mismatch between the reported NO_x (blue) and NO_x calculated from NO and NO_2 (red).

The inset in panel (c) shows a scatter plot with two populations of data highlighting this mismatch. Data with consistent reporting units is in blue, inconsistent in red and 1:1 line is in black.

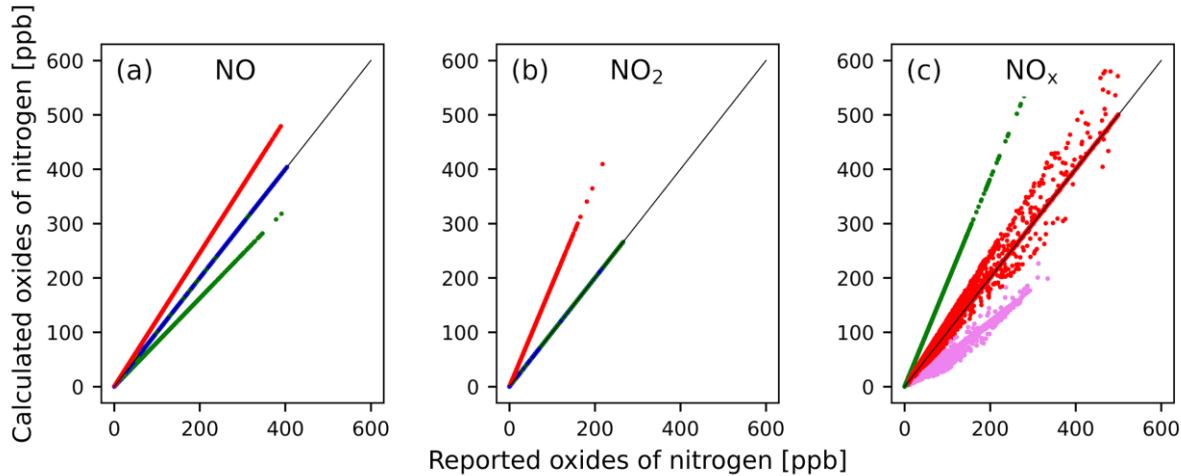


Figure 3.8: Comparison of NO, NO₂, and NO_x before and after addressing unit inconsistency issues. The figure displays the correct NO, NO₂, and NO_x data in blue, which comply with the CPCB protocol and are reported in appropriate units, and were unchanged post unit correction as it populates along 1:1 black line. The incorrect reporting of NO, NO₂, and NO_x in ppb is shown in red, with the incorrect reporting of NO_x in $\mu\text{g m}^{-3}$ shown in violet, and incorrect reporting of NO in ppb shown in green.

3.3.2 Implications of addressing the air quality data issues

Dubious air quality data has significant implications, particularly for policy making and public health. We analyse the impact of the data cleaning tool across all three years (Figures 3.10-3.12) and observe changes in annual means of NO₂ (+88.7%), O₃ (+2.3%), NO (-12.3%), NO_x (-87.9%), PM_{2.5} (-0.1%) and PM₁₀ (-0.7%) across all sites between 2019-2021.

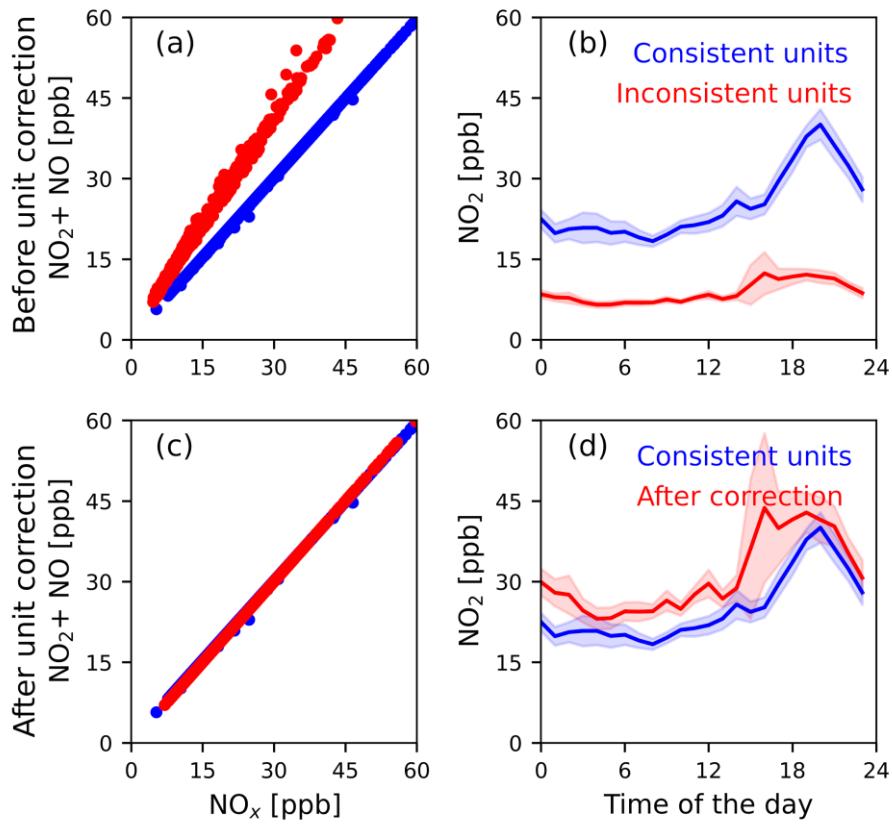


Figure 3.9: Impact of addressing unit inconsistency issue on diurnal variability of NO_2 concentrations. Panels (a) and (c) show the NO_x and $\text{NO} + \text{NO}_2$ data before and after addressing the unit inconsistency issue in NO and NO_2 , respectively, for the month of August 2020 at a monitoring site in northern India. Panels (b) and (d) show the corresponding diurnal variability of the NO_2 data associated with panels (a) and (c), respectively. The diurnal pattern of NO_2 concentrations before addressing the unit inconsistency issue (panel a) is shown in red, and after the issue was addressed (panel d), it is also shown in red.

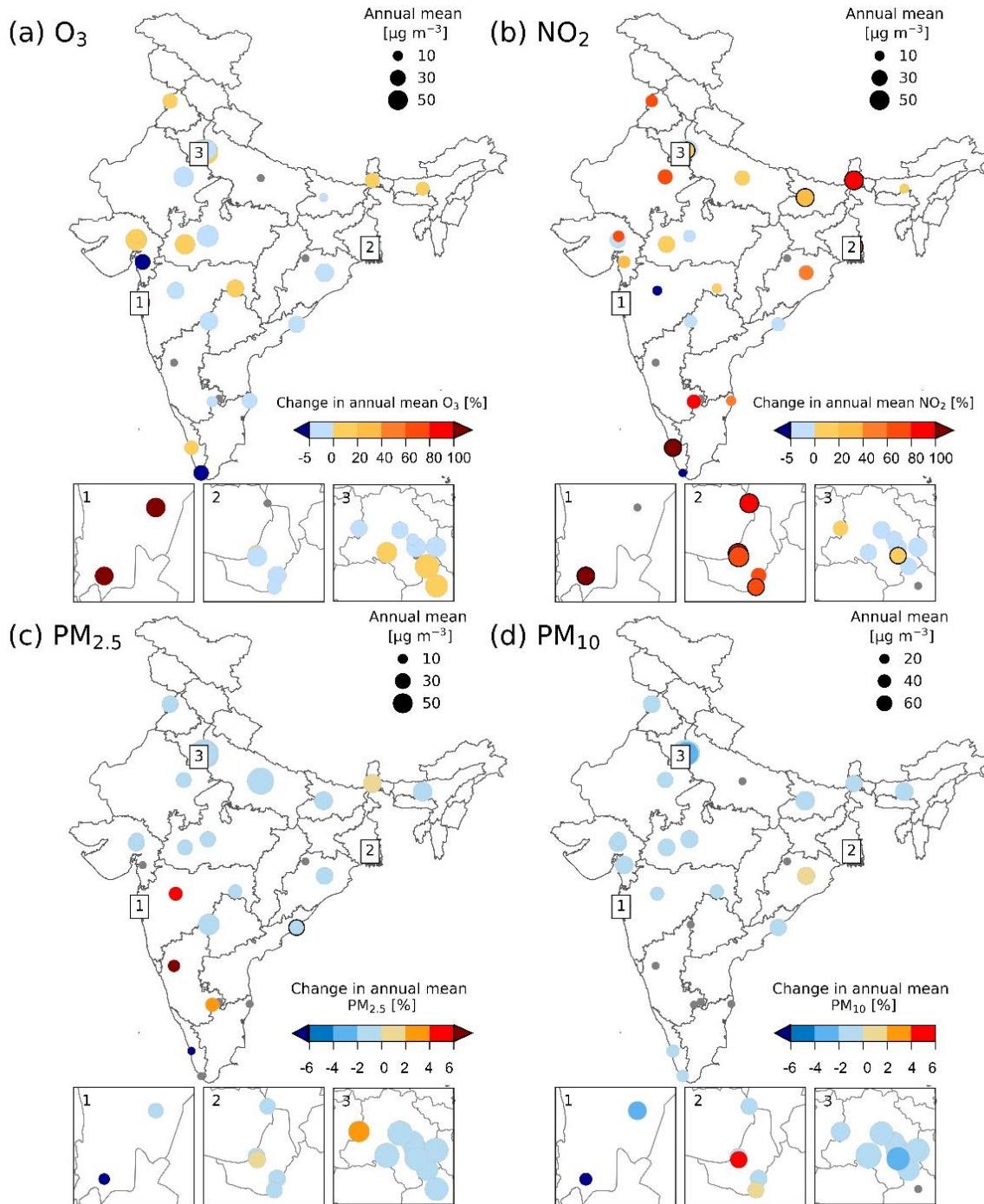


Figure 3.10. Impact of air quality data cleaning on annual means of O₃ (a), NO₂ (b), PM_{2.5} (c) and PM₁₀ (d) in 2020. The sites that showed changes in compliance with the NAAQS are indicated with a black outline. Size of data points represents the annual mean concentrations of air pollutants. The map of India shows state boundaries and the insets show district boundaries, with insets 1 for Mumbai, 2 for Kolkata, and 3 for Delhi. The annual mean

concentrations were calculated by averaging the data at a 24-hour rolling window for PM_{2.5}, PM₁₀, NO₂, and at an 8-hour rolling window for O₃. Grey data points indicate sites with less than 75% of data available post-cleaning. The state and district boundaries are from GitHub repositories (Josh Brobst, 2020; Sajjad Anwar, 2014).

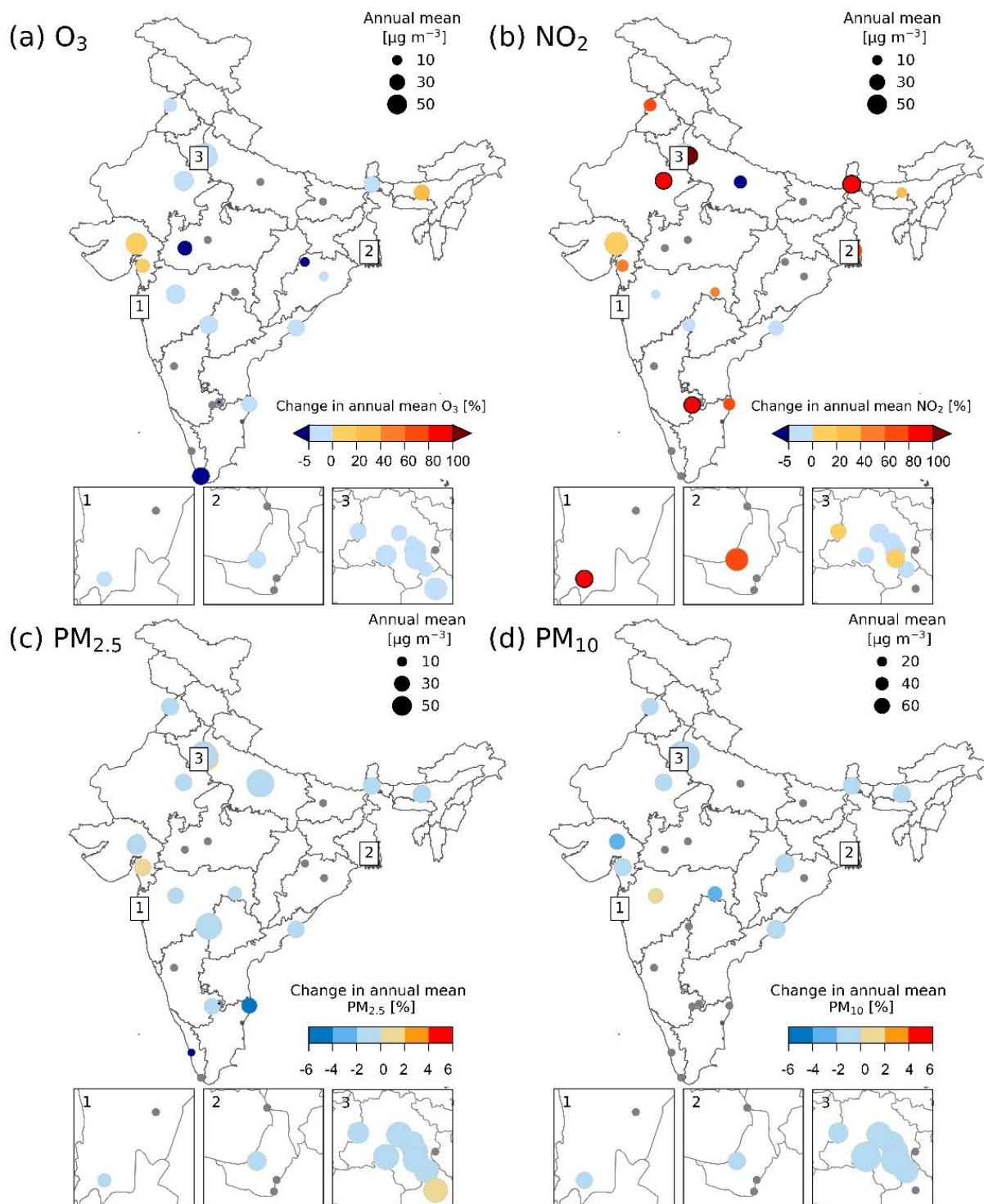


Figure 3.11: Impact of data cleaning on annual mean of O_3 (a), NO_2 (b), $PM_{2.5}$ (c), PM_{10} (d) and change in the compliance with NAAQS standards for 2019. The size of the data points in each panel is proportional to the annual mean of NO_2 after the data cleaning process, and black outlines indicate sites where there were changes in compliance with NAAQS standards. The

India map shows state boundaries and the insets show district boundaries, with inset 1 for Mumbai, 2 for Kolkata, and 3 for Delhi. Grey data points indicate sites with less than 75% data availability before or after data cleaning. The state and district boundaries are from GitHub repositories (Josh Brobst, 2020; Sajjad Anwar, 2014).

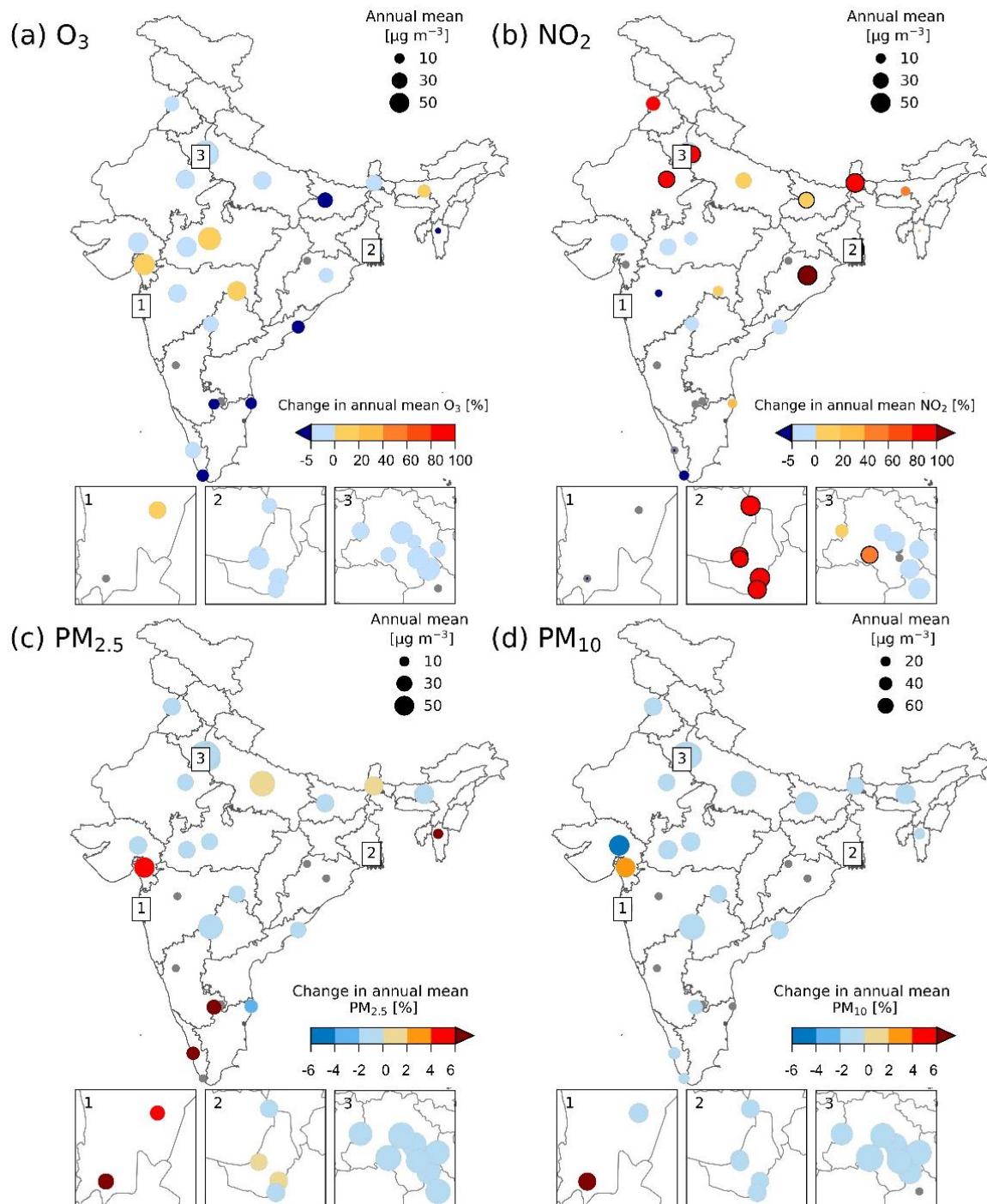


Figure 3.12: Impact of data cleaning on annual mean of O_3 (a), NO_2 (b), $PM_{2.5}$ (c), PM_{10} (d)

and change in the compliance with NAAQS standards for 2021. The size of the data points in each panel is proportional to the annual mean of NO₂ after the data cleaning process, and black outlines indicate sites where there were changes in compliance with NAAQS standards. The India map shows state boundaries and the insets show district boundaries, with inset 1 for Mumbai, 2 for Kolkata, and 3 for Delhi. Grey data points indicate sites with less than 75% data availability before or after data-cleaning. The state and district boundaries are from GitHub repositories (Josh Brobst, 2020; Sajjad Anwar, 2014).

The changes for NO₂, NO and PM₁₀ are statistically significant (p-value <0.05). Annual mean NO₂ increased by 115-522% at 21 sites in at least two study years, while O₃ increased by 5-156% at 13 sites in at least one study year after cleaning the data. Annual mean across all trace gases increased by 2.2-3.9% and PM by 0.01-0.04% after removing consecutive repeats (Figure 3.2c). After outlier removal, the annual mean of trace gases decreased by 1-4% and PM by 0.5-0.7% (Figure 3.2d). Notably, the annual mean concentration of all trace gases for a site in southern India heavily affected by radio frequency interference significantly decreased by 32-51% (p-value <0.05) on average across all three years after removing those outliers. Additionally, we found that addressing unit inconsistency led to significant increases of 88% in annual mean NO₂ and 23% in NO at 12 sites that continuously misreported both the species in ppb instead of $\mu\text{g m}^{-3}$ over the three-year period (Figure 3.2e). Such discrepancies in units, in particular for the same site, influence the long-term trends, evaluation of policy effectiveness and findings from studies such as assessment of COVID-19 lockdown on NO₂ reduction.

After data cleaning, we see a decrease of 20% on average in NO_x-to-NO₂ ratio across the study period (Figure 3.13a). The change in NO_x-to-NO₂ ratio varies from -76% to 151% and can influence the air quality research outcomes as it serves as an indicator of the degree of photochemical oxidation in the atmosphere and NO-to-NO₂ chemistry (Kimbrough et al., 2017).

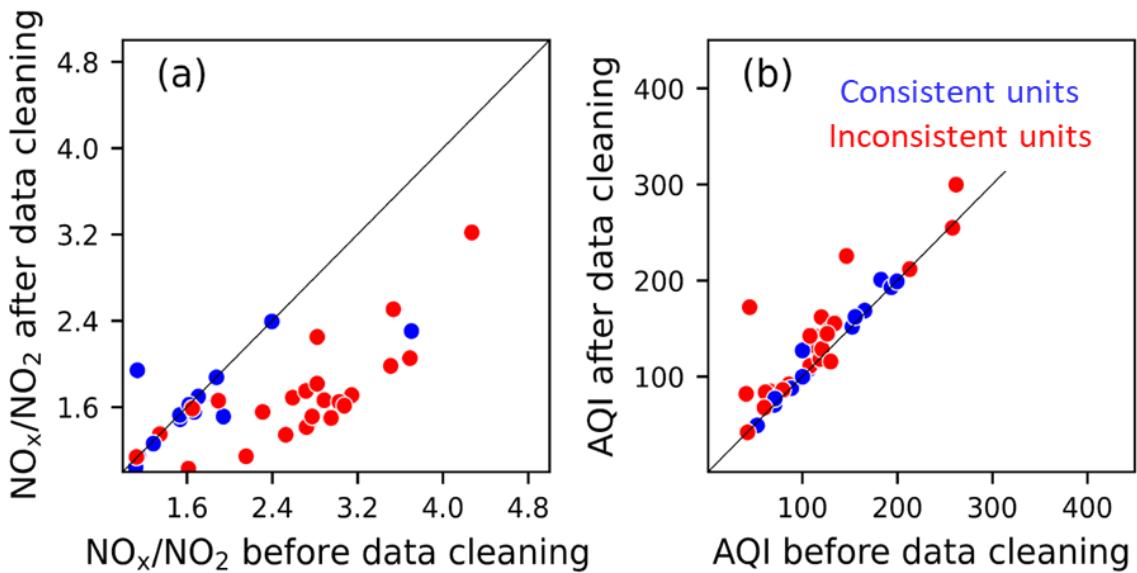


Figure 3.13. Impact of data cleaning on air quality index and NO_x to NO_2 ratio. The Figure illustrates the impact of data cleaning on the ratio of NO_x to NO_2 (in panel a) and the air quality index (in panel b) for the year 2020. The colors distinguish between the presence or absence of unit inconsistency issues in the oxides of nitrogen data and the 1:1 line is in black.

Our findings also demonstrate a significant increase of 11.4% (p -value <0.05) in the AQI, a vital index for communicating air quality to the public. We observe 21 instances with a drastic increase of 23-286% for annual AQI after data cleaning, with an average of 50% increase at 15 sites (Figure 3.13b) in at least one sample year. We find 25 instances in 15 sites (5 in 2019, 9 in 2020: and 11 in 2021: Figure 3.10-3.12) where sites become non-compliant with NAAQS NO_2 standards after data cleaning. Similarly, a site in south India became compliant with NAAQS $\text{PM}_{2.5}$ standards after data cleaning in 2020. This is concerning as it could have potentially delayed the implementation of critical interventions that are urgently needed to address air quality issues in those locations. Given the detrimental health effects, as documented in previous studies (Brown et al., 2022; Orru et al., 2013), the presence of data quality issues has significant implications for public health and reinforces the need for enhanced quality control protocols in real-time air quality monitoring.

3.3.3 Limitations of our data cleaning tool

Our tool removes recurring repeats, invalid negative records, outliers and identifies and corrects unit inconsistencies in the dataset. However, the tool cannot effectively address a series of repetitive fluctuations (Figure 3.14a) that repeat after a specified period and do not exhibit patterns that represent actual ambient conditions (Figure 3.14b).

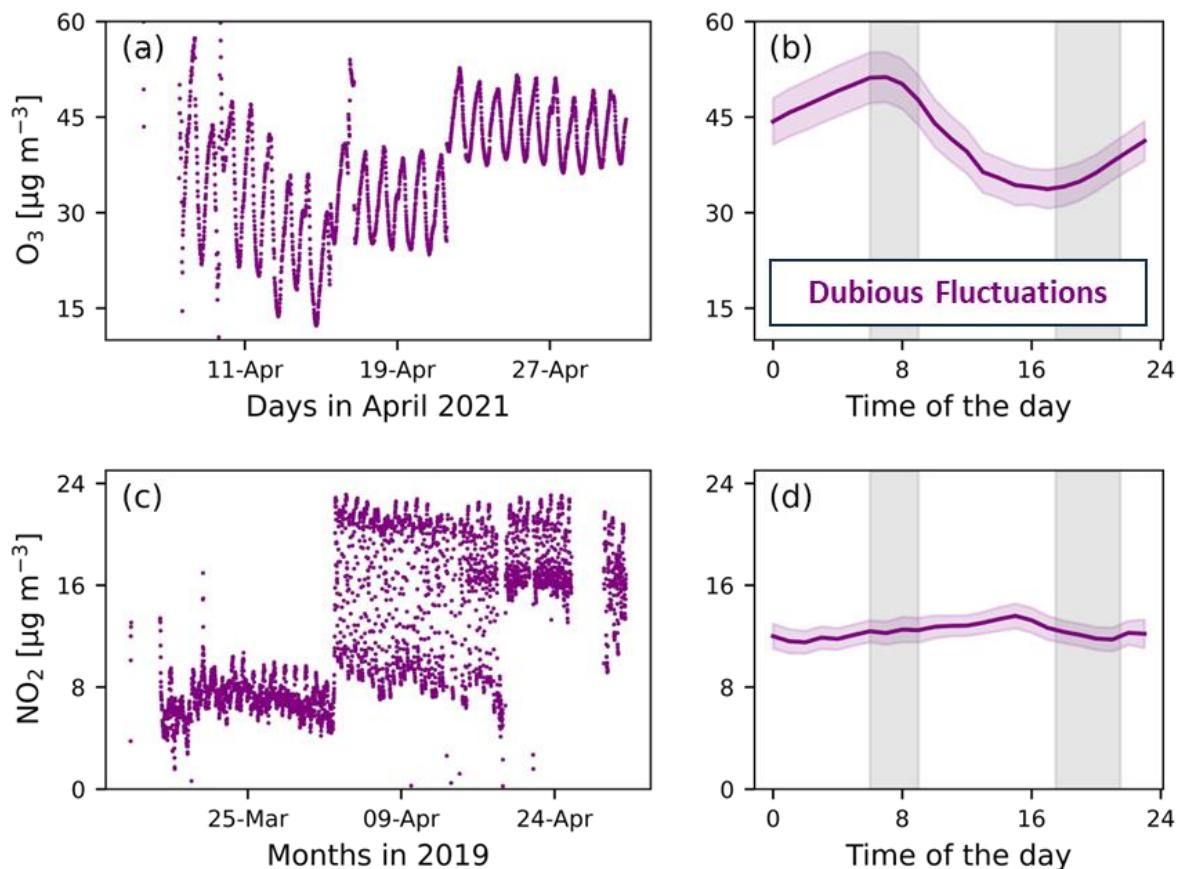


Figure 3.14. Identification of dubious waves in trace gas datasets. Panels a and c display the time series data for O_3 and NO_2 respectively. Panels b and d exhibit the corresponding diurnal variability for these datasets.

The tool cannot identify dubious data exhibiting a significant spread (Figures S14c and 3.15a), thereby being unable to flag dubious fluctuations. We also notice unexplainable constraints in the minimum and maximum concentrations for PM (Figure 3.16) at three sites. We recommend future research aimed at refining the air quality data cleaning tool.

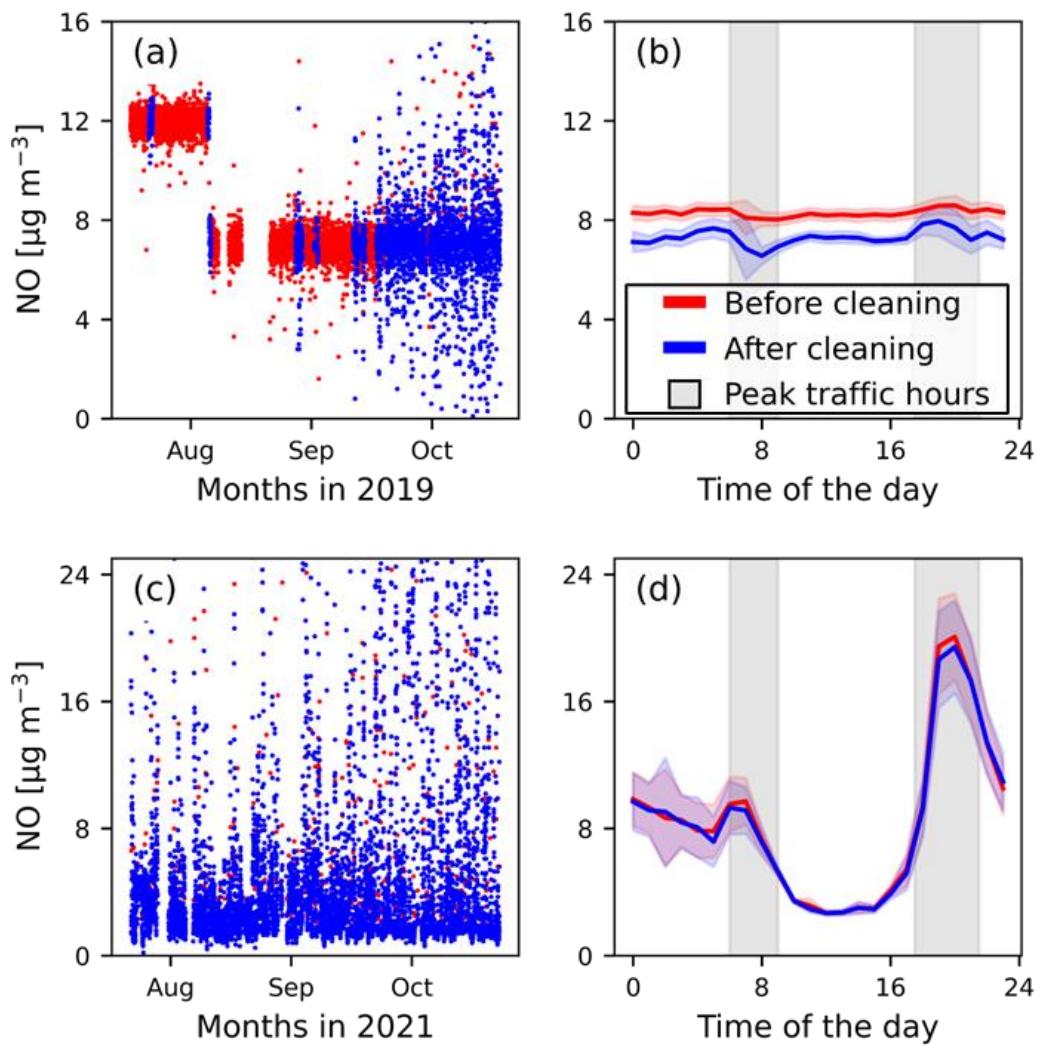


Figure 3.15. Identification of dubious data spread and partial cleaning using a data cleaning tool. Panels (a) and (c) show the NO data from a monitoring site in northern India that was corrected by removing a portion of the data (depicted in red) using the tool. Panels (b) and (d) display the corresponding diurnal variation of NO for the years 2019 and 2021, respectively. Each data point represents a 15-minute average. A comparison of the 2019 and 2021 data suggests that the 2019 data may contain faulty measurement data.

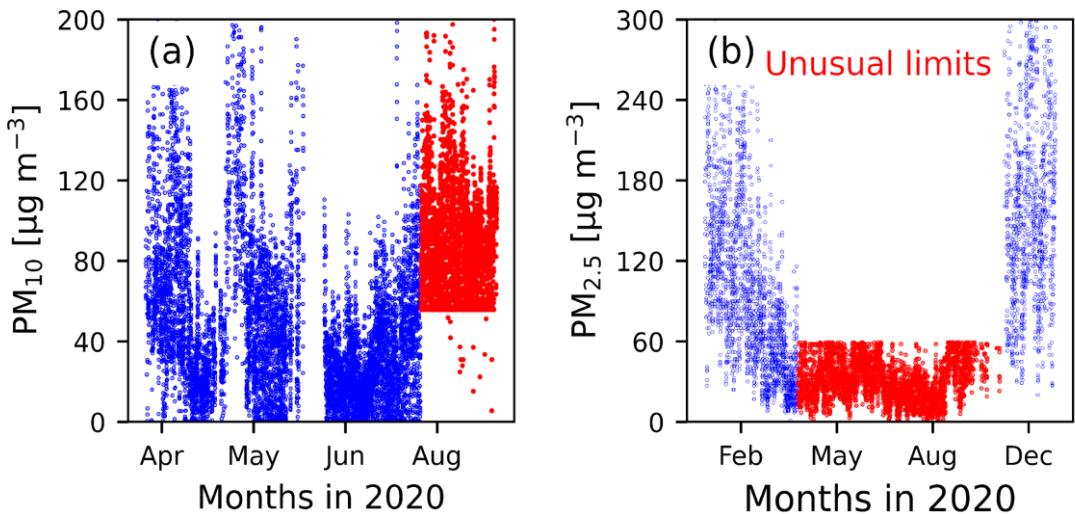


Figure 3.16. Unusual limits on PM dataset. Panel (a) displays the PM_{2.5} data with an upper limit of 59 $\mu\text{g m}^{-3}$ in red, and panel (b) shows the minimum possible value of PM₁₀ constrained at 10 $\mu\text{g m}^{-3}$ in red and other observations are given in blue. Each datapoint represents 15-minute averages.

3.4.4. Implications for future advancements in data quality management

Notable variations in air quality metrics and invalidation of 15.5% of total data after data-cleaning suggest an urgent need for data ratification. Current consecutive finding algorithms in CPCB's SMS alert system (CPCB, 2018) could be replaced with more robust metrics such as CoV. Future studies could consider incorporating mean, variance, autocorrelation, and diurnal pattern to evaluate the validity of large anomalous fluctuations. While CPCB cautions against using these records as absolute measures, it is imperative to allocate resources for proper functioning of CAAQMS in addition to instrument procurement. Urgent measures should be taken by data providers to rectify non-compliance with the parameter reporting protocol.

3.4.4. Evaluating the Impact of Data Cleaning on Air Quality Studies

In this study, we employed a data cleaning algorithm for NO₂ data collected from CAAQMS sites in Mumbai, Delhi, and Kolkata. These sites were previously utilized in several air quality studies^{83–88} that examined the impact of COVID lockdown on air pollution.

Following the established methodologies described in those studies⁸³⁻⁸⁸, we compared the reported averages with the averages obtained after applying a more robust data cleaning tool, as discussed in our study. Our findings demonstrate that data cleaning considerably affects the observed changes in pollution levels before and after the lockdown, even shifting the direction of these changes (positive /negative). Additionally, the percentage change becomes more pronounced than the most reported values in the published literature. For example, a study in Delhi reported a reduction in NO₂ levels following the lockdown. However, upon data cleaning, we observed no discernible improvement in air quality post-lockdown. Post rigorous data cleaning, we observed NO₂ levels of 78.22 µg m⁻³ in Kolkata and 60.82 µg m⁻³ in Mumbai before the lockdown, indicating an average increase of ~77% compared to prior COVID studies. Similarly, after cleaning, NO₂ levels during the lockdown were at 20.93 µg m⁻³ in Kolkata and 12.93 µg m⁻³ in Mumbai, which is ~44% more than previously reported values. These significant increases in observed NO₂ concentration are primarily found in Kolkata and Mumbai where the unit inconsistency issue is predominant (Figure 3.17). In contrast, studies that examined the lockdown's impact in Delhi using NO₂ data from all available monitoring stations were minimally affected by our data cleaning algorithm (Figure 3.17). This suggests that most sites in Delhi diligently adhere to the parameter reporting protocol of the CPCB and exhibit fewer data quality issues related to outliers and consecutive repeats. Although we did not observe a substantial difference in the occurrence of data quality issues between CPCB sites (5 sites) and non-CPCB sites (35 sites) (Figure 3.18), significant variations in data quality issues were found among Delhi, Kolkata, and Mumbai. The majority of stations in Delhi consistently followed the prescribed protocol for parameter reporting, while all stations in Kolkata and Mumbai exhibited a lack of strict adherence to the protocol and also the data consist of other data quality issues such as consecutive repeats and outliers in abundance. These noticeable differences in data quality issues between sites associated with different

organizations may be attributed to systematic factors, such as a lack of strict adherence or awareness regarding parameter reporting protocols within specific organizations.

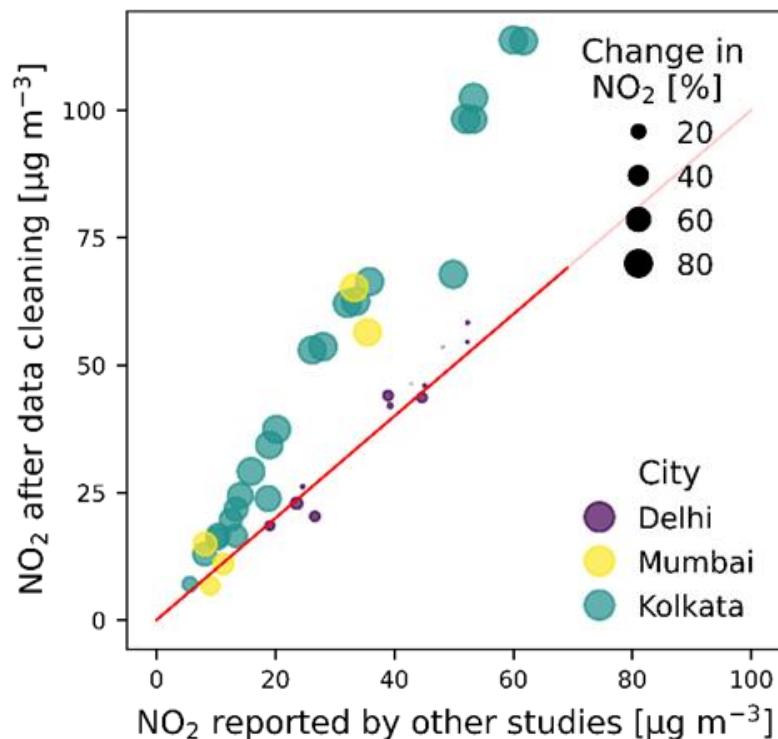


Figure 3.17. Impact of NO₂ data cleaning in published air quality studies in India. The figure illustrates the absolute NO₂ ($\mu\text{g m}^{-3}$) values reported in air quality studies^{83–88} using CAAQMS from Delhi, Mumbai, and Kolkata. The x-axis represents the absolute NO₂ values reported in the studies, while the y-axis represents the averages computed after data cleaning in the current study. The size of each data point is proportional to the change in NO₂ values observed after data cleaning, relative to the absolute values reported by the CAAQMS sites.

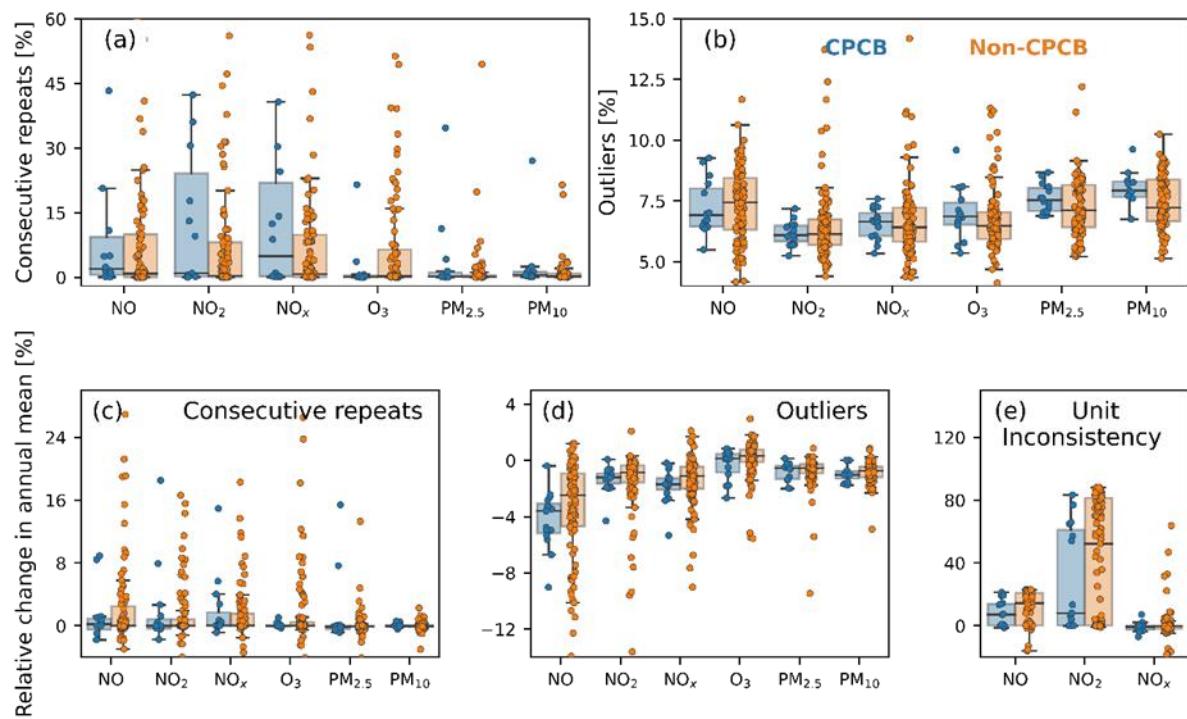


Figure 3.18. Percentage of invalid data and its influence on annual mean pollutant levels (2019-2021): Dubious data identification within CPCB and Non-CPCB sites CAAQMS data. Panels (a-b) display the percentage of consecutive repeats (a) and outliers (b) removed from the total available data for each site from 2019 to 2021. Panels (c-e) illustrate the relative change in annual mean after the removal of consecutive repeats (c), outliers (d), and ensuring unit consistency (e). Bar colours distinguish data from different authorities: CPCB in blue, non-CPCB (SPCBs, IITM, IMD) in orange, with each data point inside the bar representing a specific site in each year (2019-2020).

3.4.5. Code availability

The Python scripts used for data cleaning and analysis are available at the <https://github.com/Madhumitha11-s/AirPy.git>

Chapter 4

Preliminary Exploratory Data Analysis

4.1 Satellite Based Ground Level Estimates:

Air pollution is a result of complex chemistry involving multiple factors such as emission, meteorology such as rain, humidity, wind direction and background chemistry, and long-range transport. The comparison with absolute concentrations without considering the meteorological parameters, might lead to misleading interpretations. We aim to utilize satellite-based ground-level estimations to quantify the impact of lockdown on air pollution. In order to derive strong conclusions from it, initial sanity checks on these models were performed along with the COVID impact analysis. Both CPCB station data and satellite-based measurements were checked to evaluate whether the variability at the monthly level was reproduced or not.

4.1.1 Spatial Variation in Particulate Matter Response due to Lockdown:

The PM_{2.5} satellite-based ground level estimations were available for the year 2019 and 2020 at monthly level with a spatial resolution of $0.01^\circ \times 0.01^\circ$ ($1.11\text{ km} \times 1.11\text{ km}$) and $0.1^\circ \times 0.1^\circ$ ($11.1\text{ km} \times 11.1\text{ km}$), whereas for NO₂ data was only available at $0.01^\circ \times 0.01^\circ$ for the first six months of 2019 and 2020. Data visualizations of these data were performed. Fig 4.1 shows that PM_{2.5} in the April month of 2020 was high when compared to pre-COVID year. The lockdown started on March 25, 2020 and extended till April 14 throughout the nation, and there was no significant impact on PM_{2.5} due to the lockdown in the month. Notably, the PM_{2.5} in February and April has increased significantly in central Mumbai and the outskirts of Mumbai (Fig 4.2). Fig 4.6 shows no statistically significant decrease in PM_{2.5} surface estimated during the lockdown month. Also, the increase in PM_{2.5} in the outskirts of the MMR region was unexplainable by anthropogenic emissions. The Sentinel Land Cover Data showed that Scrubs, grasses, and bare grounds surround the outskirts of MMR for more than 90%. Due to huge built-up areas, the PM_{2.5} increase in Central India could be attributed to anthropogenic

emissions. A study in a coastal Industrial area, Trombay, Mumbai, demonstrated that only 22 % of fuel and oil combustion contributed to PM_{2.5}⁸⁹, which might not be sufficient to significantly reduce PM_{2.5} during lockdown. The increase could be attributed to the direct PM_{2.5} emission or secondary formation of PM_{2.5} from precursor gases emitted in the farm ⁹⁰. The uncertainties in PM_{2.5} are still large due to aerosol chemistry's complex and synergetic effects.

There are multiple driving factors for PM_{2.5} accumulation, including emissions, unfavourable meteorology, secondary formation of nitrate, sulphate, and organics. The pollution reduction strategies must also consider meteorological factors to determine the effect of the controlling procedures. The response of PM_{2.5} during COVID lockdown is not straightforward due to the complex chemical reactions between precursors and hydroxyl radicals (OH) and peroxy radical intermediates (HO₂ and RO₂).

PM_{2.5} accumulated near the coastal areas increases between Oct to Feb months. Mumbai experienced much calmer wind conditions in Oct (0.75 m s⁻¹) and Nov (1.32 m s⁻¹) and until Apr (1.3 - 1.4 m s⁻¹) (Table 4.1). In Dec - Feb wind is blowing along with the coastal areas from north to south. More stagnant conditions lead to increased PM_{2.5}. The land-sea breeze interaction during other months might have also prevented the satellite from capturing the plume well. Therefore, it is possible that the satellite captures the plumes when the wind is just along the coast but not when it is across.

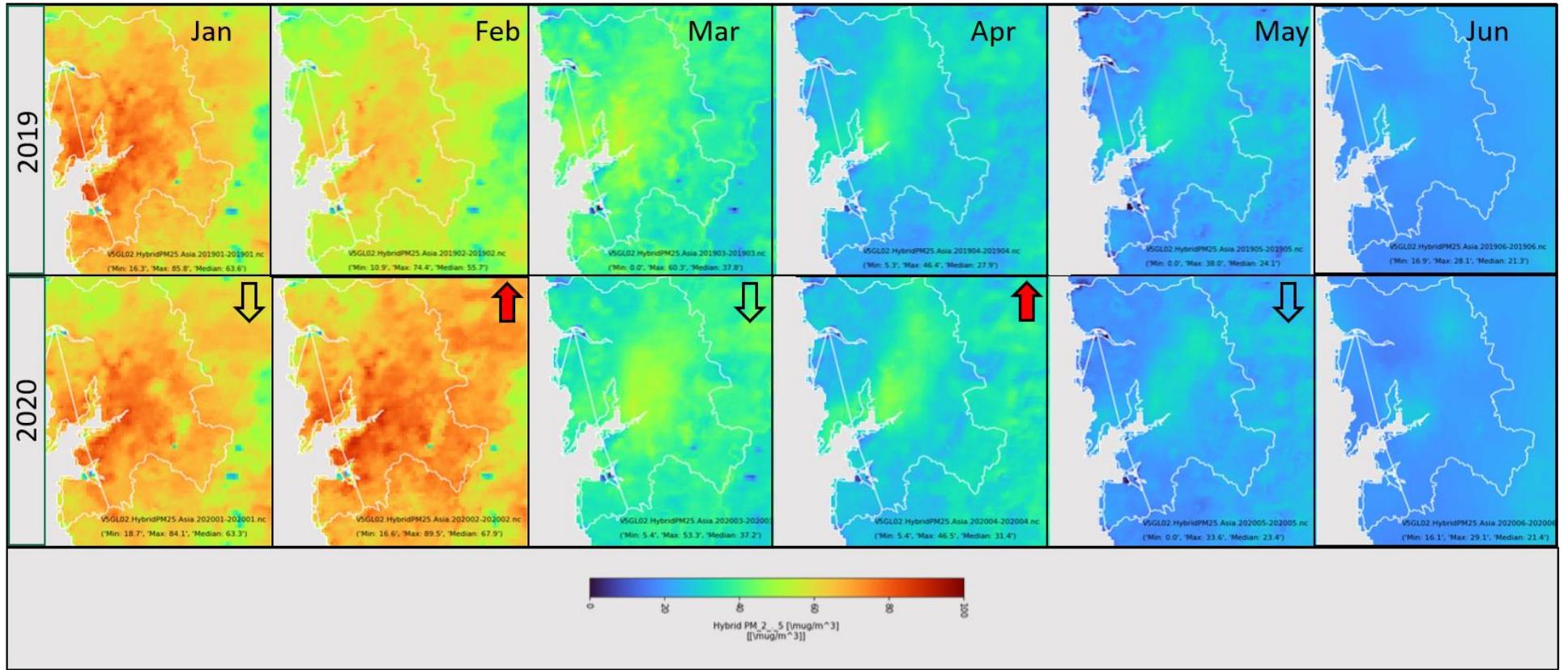


Figure 4.1: Monthly mean PM_{2.5} ($\mu\text{g}/\text{m}^3$) in $0.01^\circ \times 0.01^\circ$ for year 2019 and 2020 (Jan – Jun). The arrows in 2020 visualizations indicates the increase or decrease in PM_{2.5} in 2020 with respect to 2019 (Red color arrow – indicating the increase). Annotations inside each subplots has descriptive statistics such a minimum, maximum, standard deviation, mean concentration for values within MMR

Table 4.1: Wind speed in MMR region from 2018-2019. Beaufort scale was used for distinguishing high wind to calmer wind. Calm: $<0.5 \text{ ms}^{-1}$, light air – $0.5\text{-}1.5 \text{ ms}^{-1}$, light breeze – $1.6\text{-}3.3 \text{ ms}^{-1}$, gentle breeze - $3.4\text{-}5.5 \text{ ms}^{-1}$, moderate breeze – $5.5\text{-}7.9 \text{ ms}^{-1}$, strong breeze $> 10.8 \text{ ms}^{-1}$.

Wind speed (m s^{-1}).	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2021	1.47	1.23	1.45	1.22	1.48	2.70	3.60	2.48	2.81	1.03	1.47	1.38
2020	1.35	1.33	1.31	1.24	2.00	2.38	2.85	4.27	0.79	0.25	1.58	1.46
2019	1.51	1.43	1.40	1.41	1.84	2.30	3.63	3.53	2.83	0.99	1.09	1.37
2018	1.27	1.24	1.07	1.35	2.16	3.10	4.60	4.40	1.29	0.73	1.17	1.53
Average	1.40	1.31	1.31	1.31	1.87	2.62	3.67	3.67	1.93	0.75	1.33	1.43

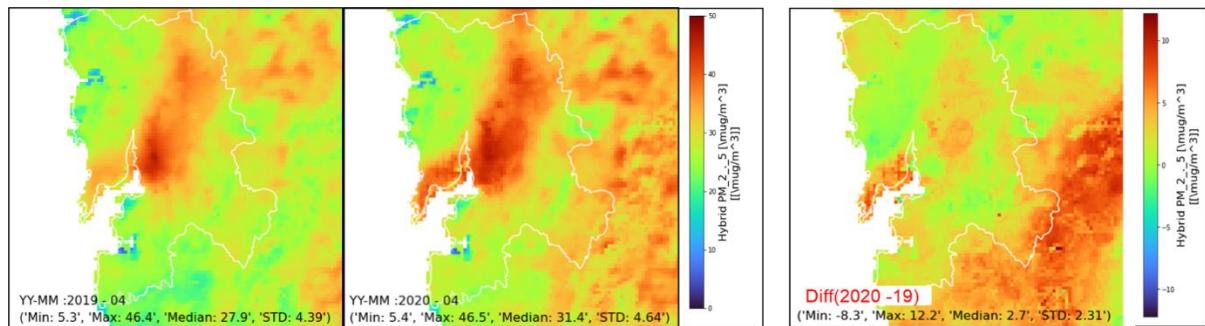


Figure 4.2. April month PM_{2.5} ($\mu\text{g}/\text{m}^3$) in $0.01^\circ \times 0.01^\circ$ for years 2019 and 2020 and the absolute increase in PM_{2.5} in 2019, the outskirts of MMR with respect to 2020. Annotations inside each subplot have descriptive statistics such a minimum, maximum, standard deviation, and mean concentration for values within MMR.

The methodology to control air pollution continues to be challenging due to the complex phenomenon of atmospheric chemistry, meteorology, and pollutant sources. The long-term goal of our project not only focused on evaluating the impact of short-term lockdowns on air quality but also to examine whether short term, partial lockdowns can be a mitigation measure when the air quality becomes problematic. We also aim to demarcate the components of meteorology and its impact during LD to measure the true effect of LD measures.

The secondary aerosol formation can also potentially complement the reduction in direct emission. Therefore, utilizing LD as a pollution control measure must be more skeptical about being successful and it is also unsustainable from the economic perspective. Stringent lockdowns extending for months can also severely affect the nation's economy. Estimates suggest that individuals' income in India dropped by 40% in the first stringent lockdown during April and May 2020, and the GDP continued to be 15% lower even till the end of 2020⁹¹.

4.1.2 Decrease in NO₂ during Lockdown & NO₂ Hotspots in MMR:

There was a substantial reduction in NO₂ in March 2020, despite the relatively high concentration in February 2020 compared with 2019 (Figure 4.4). Due to restricted vehicular movements, the NO₂ statistically reduced in April 2020 compared to April 2019. Oxidation of NO during combustion forms NO₂, whose life span is less than two days in the atmosphere (reference?). Due to the very short lifespan of NO₂, the NO₂ concentrations primarily depend on the local NO_x emissions. Global studies have shown reductions in NO₂ levels due to reduced emissions of NO_x from traffic related emissions^{92–94}.

We noticed two hot spots for NO₂ in the MMR region during our preliminary assessments. Out of which, one hotspot (MMR central hotspot) was found to be very persistent. It was noticed even during lockdown months and other hotspots in central Mumbai were not persistent during lockdown months and a reduction in pollutant intensity was noticed. The Sentinel Land Cover Data showed that the persistent hotspot of MMR contains Scrubs, grasses and bare grounds and no anthropogenic emissions are possible from that land cover. Hence, the persistent hotspot can be either due to meteorology or the transport of pollutants from the source. Figure 4.3 shows the Mumbai Central hotspot locations, which consist mainly of built-up urban areas and forest swamps. They are mainly covered by forest swamps in the south and surrounded by built-up areas. MMR central hotspots are mainly composed of barren land and forest areas. However, it is covered by built-up on the west and large agricultural cropland on the east. No ground

stations are available near the MMR central hotspot, and the stations of Airoli (40 km away) and Mahape CPCB (30 km away) are the closest sites to this hotspot.

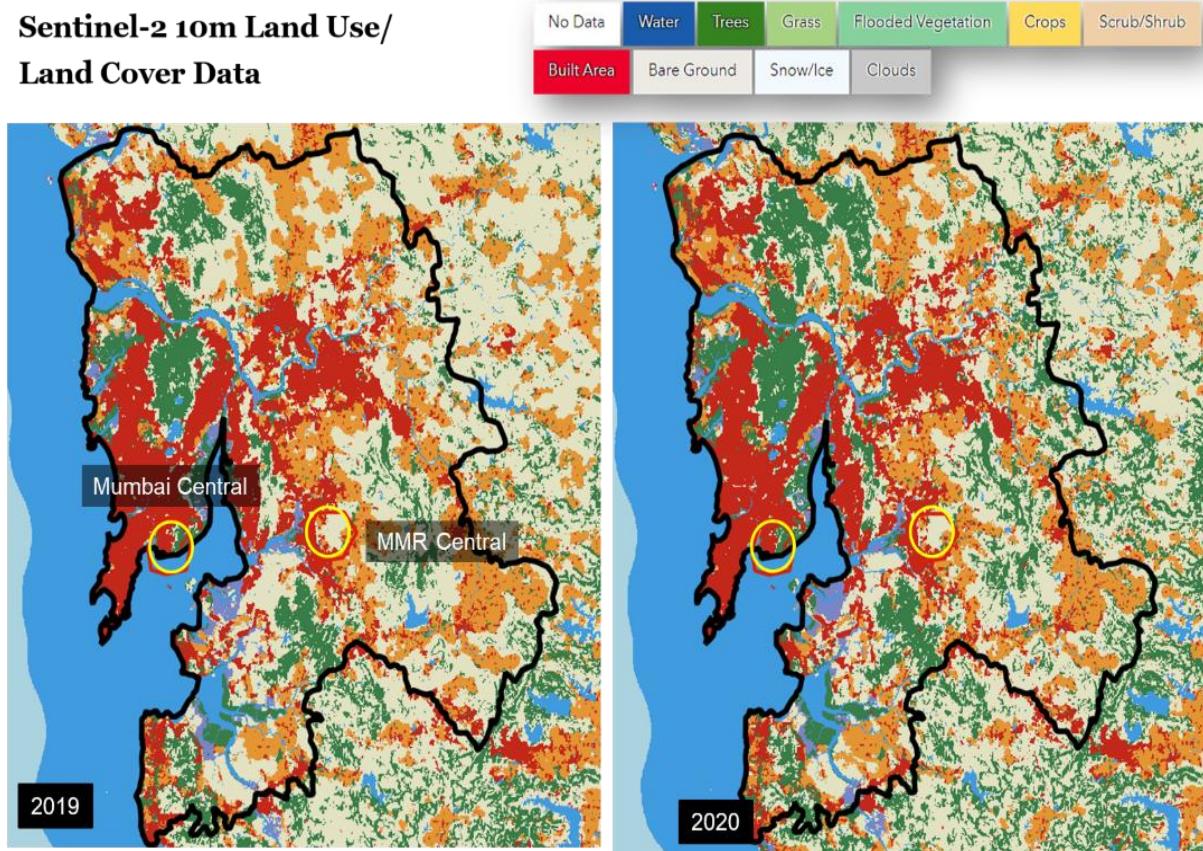


Figure 4.3. Land cover composition of MMR region in 2020 and 2019. MMR central hotspot is covered with barren land whereas Mumbai central hotspot is composed of built-up area.

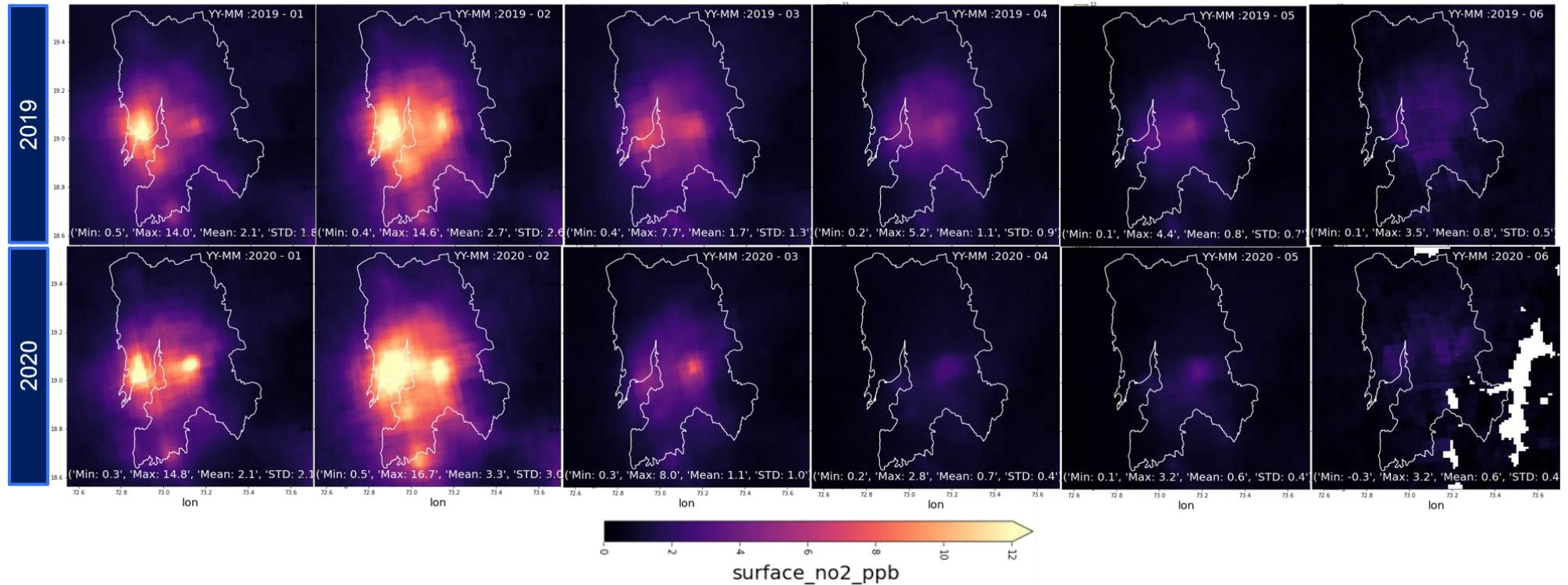


Figure 4.4: Monthly mean NO₂ (ppb) in $0.01^\circ \times 0.01^\circ$ for the years 2019 and 2020 (Jan – Jun). Concentration of NO₂ is in ppb. Annotations inside each subplot have descriptive statistics such a minimum, maximum, standard deviation, mean concentration for values within MMR.

In April of 2019 (Fig 4.5), the wind is predominantly from the north, therefore we would expect the plume to be slightly right downward. The built-up area near the MMR central could have shifted to the right to form the second hotspot.

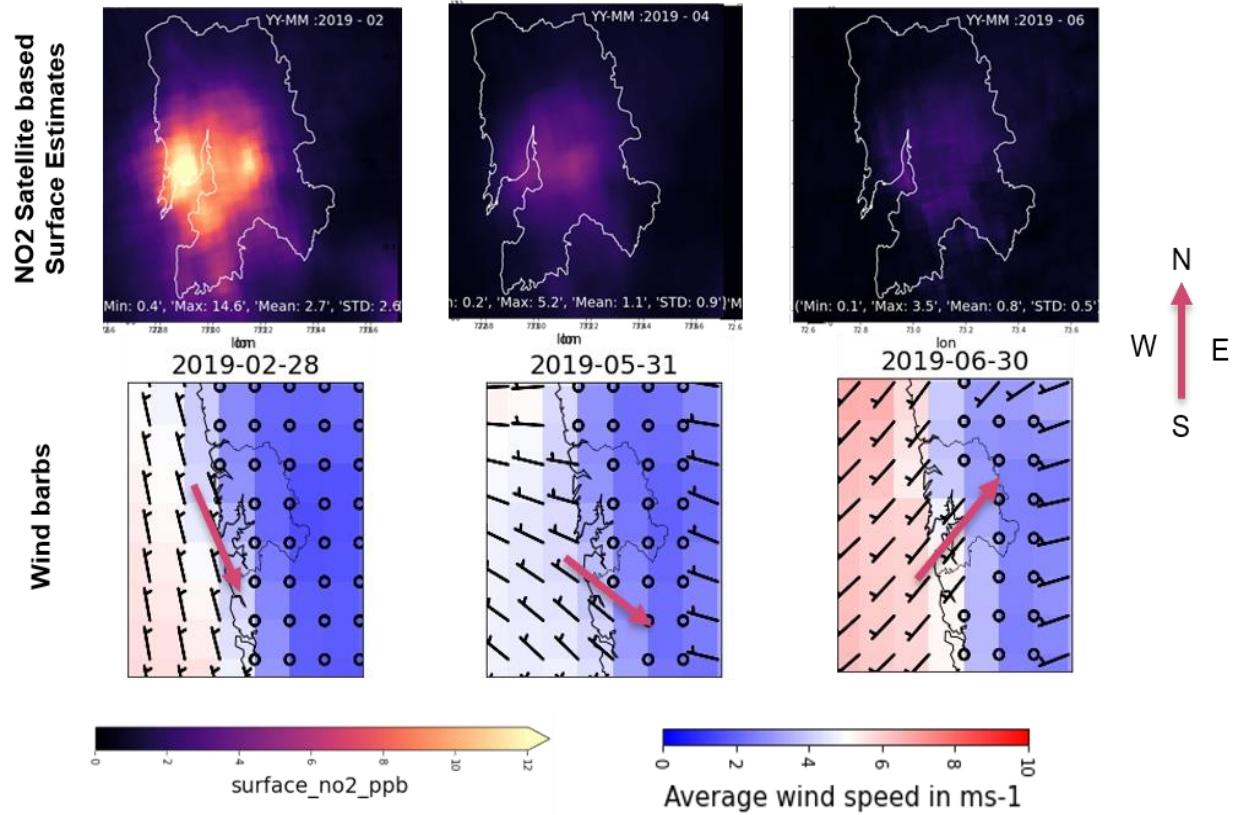


Figure 4.5: Overview on impact of wind speed on NO₂ hotspots. Panel 1 data has NO₂ satellite estimates and Panel two indicates the wind speed and direction. Calm wind condition is denoted by circles and wind barbs denote the direction of wind (For simplicity, pink arrow denotes the wind direction).

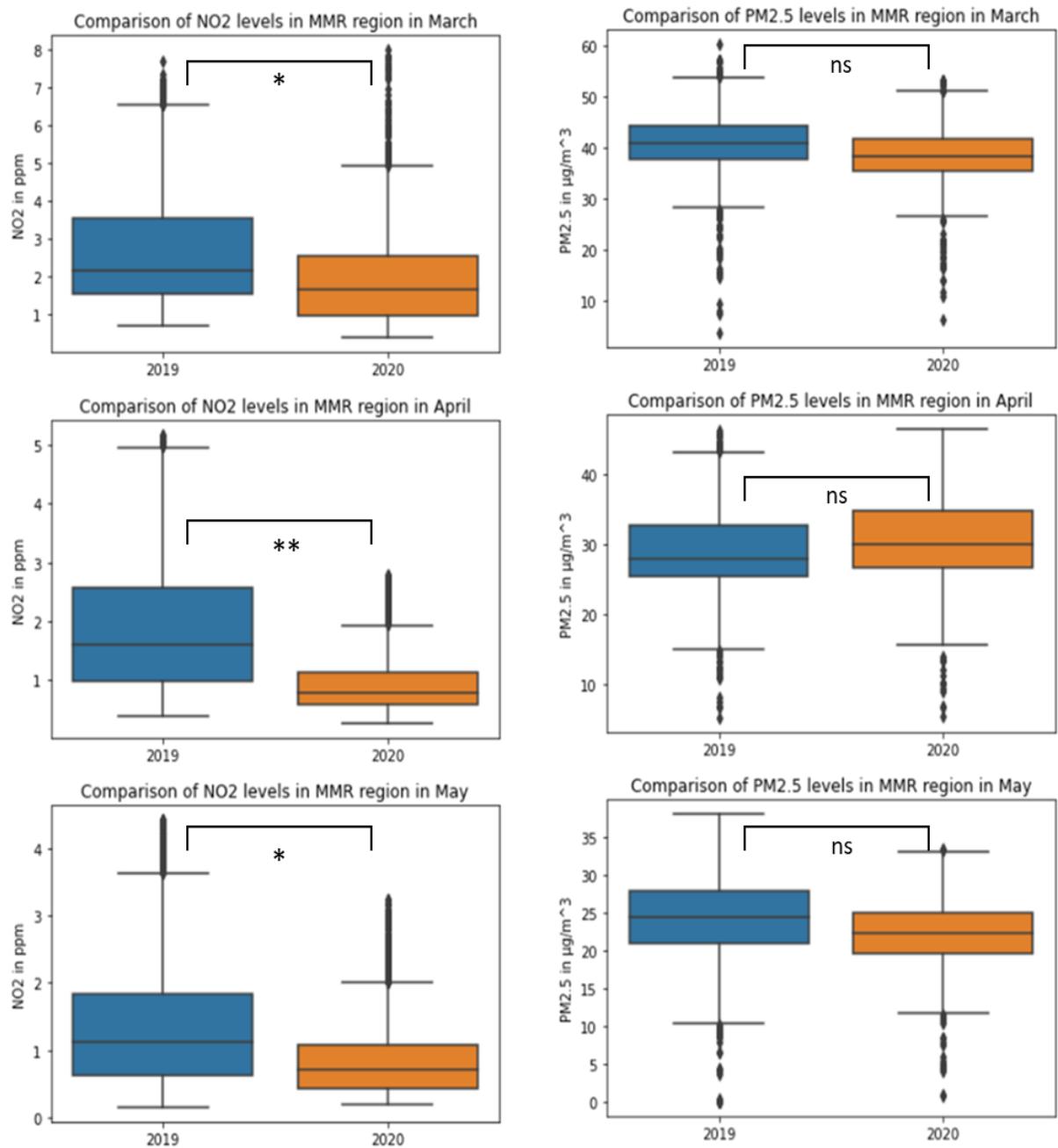


Figure 4.6: Response of NO₂ and PM_{2.5} satellite estimates during 2020 (COVID year) and 2019 (March – Pre COVID month; April, May – COVID Month). Annotations: ns: $p \leq 1$; *: $1e-2 < p \leq 5e-2$; **: $1e-3 < p \leq 1e-2$; ***: $1e-4 < p \leq 1e-3$; ****: $p \leq 1e-4$.

4.1.3 Increase in Tropospheric O₃ during Lockdown

We further analyzed the raw tropospheric column concentrations of O₃ from the OMI/ Aura instrument in 0.25° × 0.25° resolution to assess long-term trend. The annual average concentration of the O₃ total column showed an unusual reduction in 2019. The annual average showed a significant increase (Fig 3.7) in the COVID year relative to pre-COVID years, and the annual average consistently decreased after lockdown. O₃ concentrations depend on various factors such as precursor availability, season, meteorological parameters, and surrounding environment. Globally, the O₃ responses from COVID vary based on the chemical regime of each location, which is driven by the ratio of volatile organic compounds to NO_x ratio. The insolation and increased temperature can also be a potential reason for increased O₃ concentrations ⁹⁵. Reduction of NO₂, NO, NO₂ and CO during lockdown in turn, reduced the O₃ depletion due to the NO titration cycle ⁹⁶.

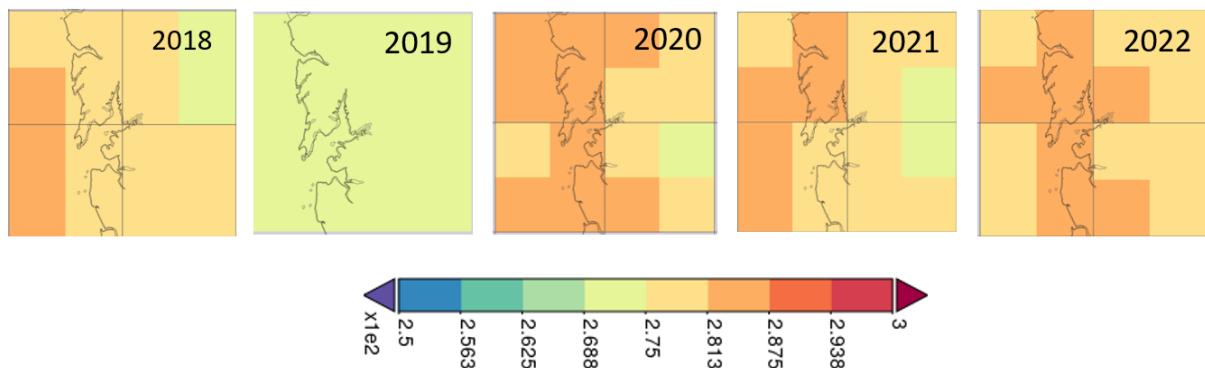
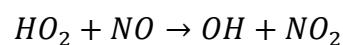
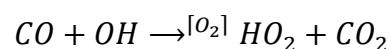
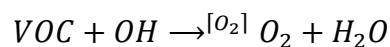
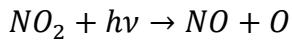


Figure 4.7: April O₃ Data comparison using Satellite measurement: OMI/Aura O₃ (O₃) DOAS Total Column Level 3 data 0.25° × 0.25°. Concentration in DU (Total column). Source: Giovanni





4.1.4 Reduction in SO₂ surface concentration estimates during Lockdown

SO₂ is the major indicator of air pollution from combustion sources and chemical fuel emissions. They function as a major precursor in the formation of nucleation in the atmosphere. The SO₂ monthly average from MERRA-2 ENSEMBLE Model showed a drastic decrease in SO₂ concentration in the lockdown month (Fig 4.8), since the commercial and incineration activities were closed as a part of pandemic control actions ⁹⁷.

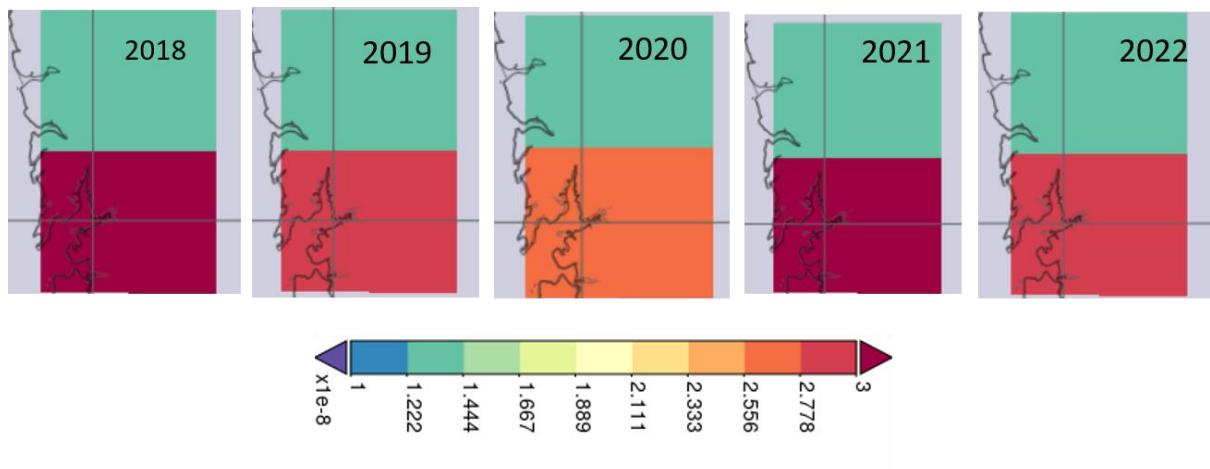


Figure 4.8: Time Averaged Map of SO₂ Surface Mass Concentration (ENSEMBLE) for April months 0.5°x 0.625°. [MERRA-2 Model M2TMNXAER v5.12.4] Concentration in kg m⁻³.

As stated earlier, air pollution mitigation is a complex problem to solve, since reducing primary pollutants might lead to temporary improvement in air quality but could aggravate the secondary aerosol formation and O₃ formation. Therefore, secondary pollution formation without primary pollutant emission should be studied to implement short term lockdowns as air quality mitigation measures.

4.2 Comparison of Satellite estimates with ground-based measurements

In order to derive strong conclusions from the satellite products, we compared ground-based and satellite measurements for our study location for lockdown and pre lockdown year. The data for comparison was extracted according to the methodology set in chapter 2 in section 2.1. The CPCB data during satellite overpass time (1:30 PM) in the case of NO₂ and the monthly mean was utilized in the case of PM_{2.5} comparison. The satellite data extracted for all stations were compared with the monthly mean to check the trend reproducibility of satellite products. Fig demonstrates the agreement in the trend variation between satellite estimates of PM_{2.5} and ground-based measurements of PM_{2.5} for Bandra MPCB site for 2020 annual year. Even though the satellite product captured the trend (Fig 4.9 & Fig 4.10), the PM_{2.5} was still biased higher and needs a correction with ground-based estimates. No negative slope was found between the estimated PM_{2.5} and ground-based measurement and demonstrated a good fit with R > 0.8 for all the stations.

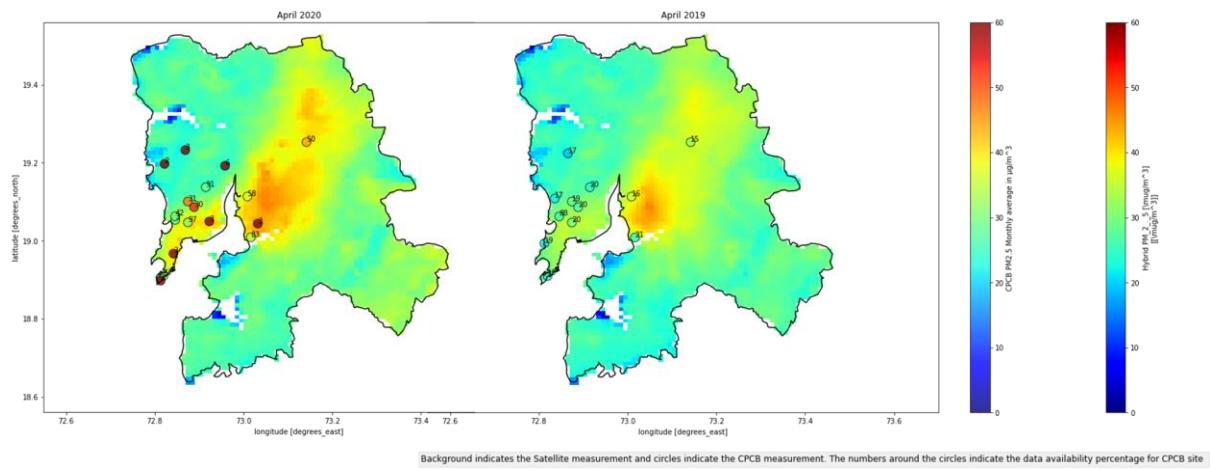


Figure 4.9: Satellite estimations (After MPCB data correction using temporal correction factor) and surface measurements for PM_{2.5} for April 2020 & 2019. The circular markers denote the CPCB measurements, and the spatial data is from satellite estimates. The annotations over the circular marker denote the data availability for the CPCB data for that month.

The NO₂ values were found to be biased lower than actual and the satellite product did not capture the trend displayed by CPCB data. These disagreements with irreproducible trends have further increased the uncertainties concerning the hotspots found in the MMR region (Fig 4.10).

Table 4.2: Sites available for long term trend studies in Maharashtra

Location	Type	Data available
MIDC, Chandrapur	Industrial	2017 - 2022
Chandrapur	Industrial	2016 - 2022
MIDC Walunj, Aurangabad	Industrial	2018 - 2022
Civil lines, Nagpur	Residential	2017 - 2022
KTHM College Campus (Gangapur Road), Nashik	Residential	2017 - 2022
Shree Pimpleshwar Mandir, Dombivli east	Industrial	2017 - 2022
Karve Road - CAAQMS, Pune	Residential	2016 - 2022
Airoli fire station, Navi Mumbai	Rural & other Areas	2017 - 2022
Municipal Corporation Premises, Solapur	Residential	2018 -2022

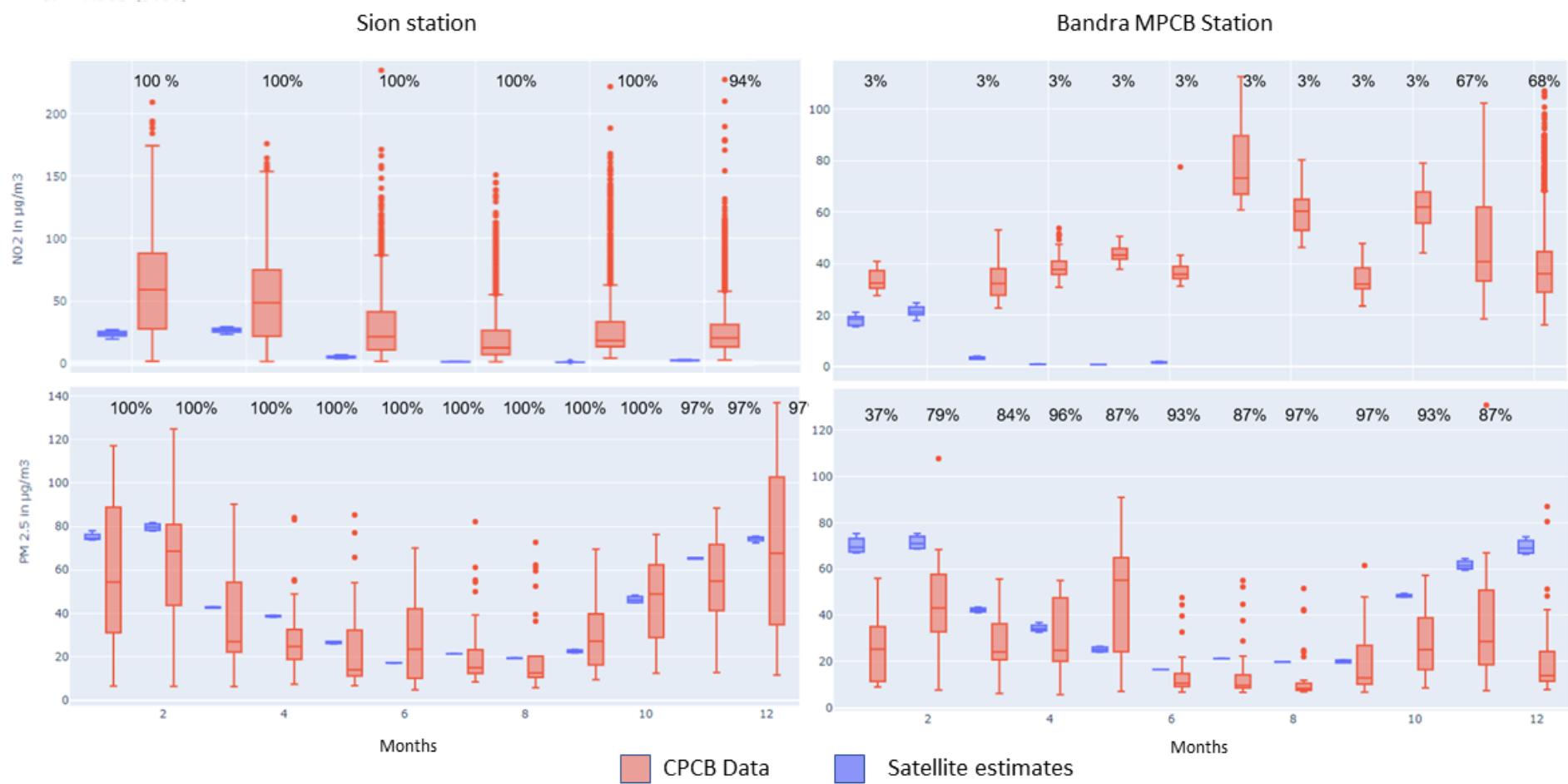


Figure 4.10: Comparison of satellite ground-based estimates with CPCB data in satellite overpass time at a monthly level. The annotations above each box denote the percentage CPCB data available for comparison.

4.3 Long term trends in ground station measurements

We extended our study outside the MMR region to study long-term trends on ground station measures in comparison with COVID year since only two sites within MMR region reported data for post COVID years. We considered four sites situated in Industrial and Residential locations (Table 4.2). The PM₁₀ values consistently decreased in all stations during April of COVID year concerning baseline (Average of 2017 – 2019). Industrial sites showed a relatively higher reduction in PM₁₀, more than 2-fold, in comparison with residential sites. Interestingly, PM₁₀ has increased in the post COVID years in industrial sites relatively higher than pre COVID years despite a stringent anthropogenic switch in 2020 (Figure 4.11). This essentially explains that lockdown could only be a short-term mitigation plan. Residential areas in Pune continue to show relatively lower PM₁₀ concentrations, even after unlock, and remain lower than pre COVID years. The response in PM₁₀ varies highly with the type of location. The trend of PM_{2.5} concentration is also similar to PM₁₀ in all the sites. In all industrial sites, NO₂ concentration has dramatically increased post COVID lockdown and no significant decrease in NO₂ was found in these locations (Fig 4.12). In contrast, Residential areas showed decrease in NO₂ during COVID and regained the NO₂ similar to Pre COVID in the post COVID years.

In Nagpur and Chandrapur Station, O₃ has increased during lockdown year compared to Pre and Post COVID years. This O₃ increase could have been due to reduced NO_x emission and reduced the O₃ depletion due to NO titration cycle ^{96,98}. Interestingly, Shree Pimpleshwar Mandir, Dombivli east station showed a 3-fold increase in O₃ post COVID lock years relative to PrL and L years. The O₃ concentrations in Nashik residential site gradually increase even after the COVID lockdown. With the limited data availability in 2019, Figure 4.15 demonstrates that the O₃ concentration continues to increase from the lockdown year for sites within MMR.

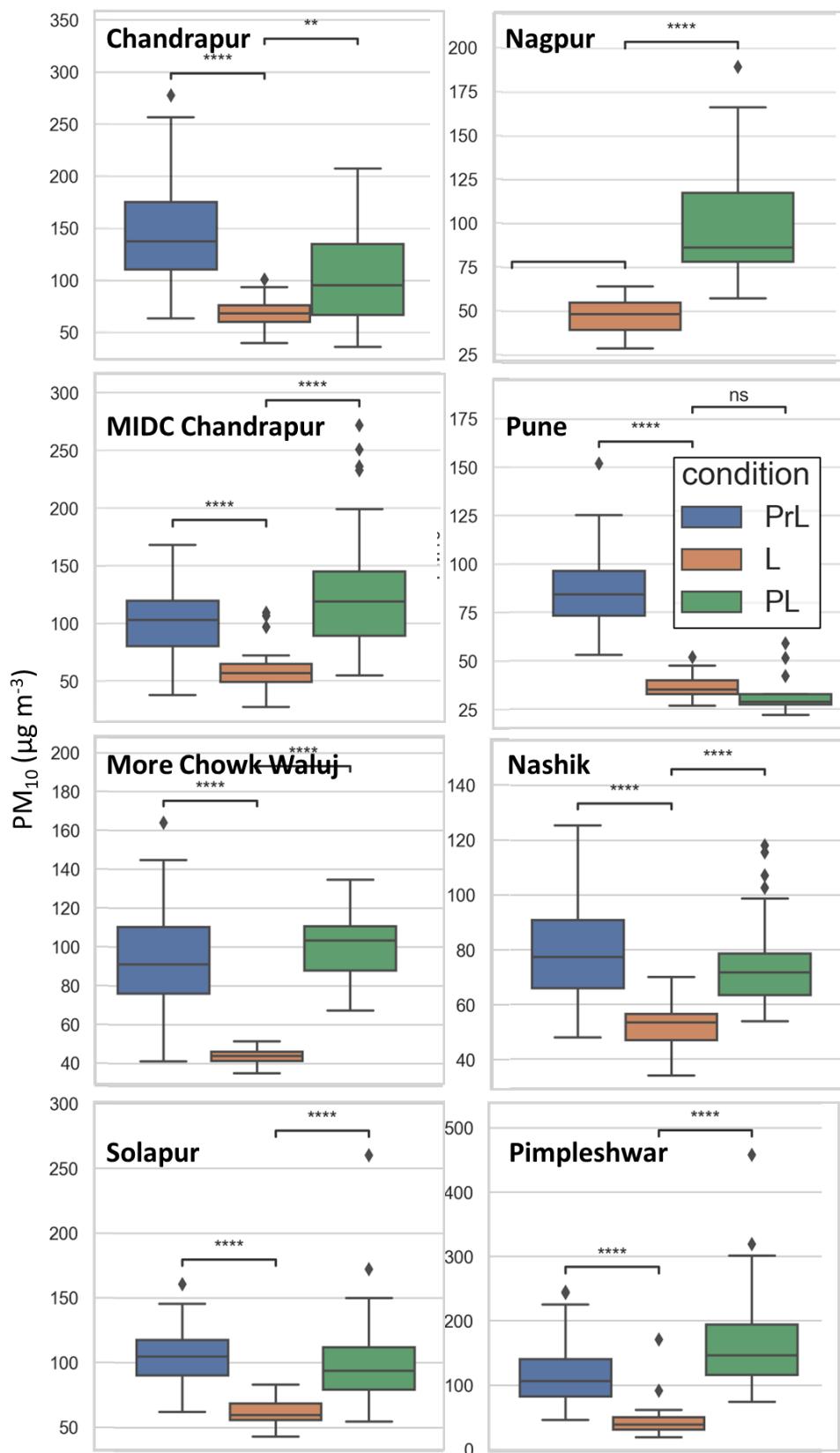


Figure 4.11: Comparison PM₁₀ during lockdown month (April). Pre lockdown years (PrL) 2017 - 2019, Lockdown (L) – 2020, Post Lockdown (PL) – 2021 -2022. (Concentration of PM₁₀ is

in $\mu\text{g}/\text{m}^3$). Annotations: ns: $p \leq 1$; *: $1e-2 < p \leq 5e-2$; **: $1e-3 < p \leq 1e-2$; ***: $1e-4 < p \leq 1e-3$; ****: $p \leq 1e-4$.

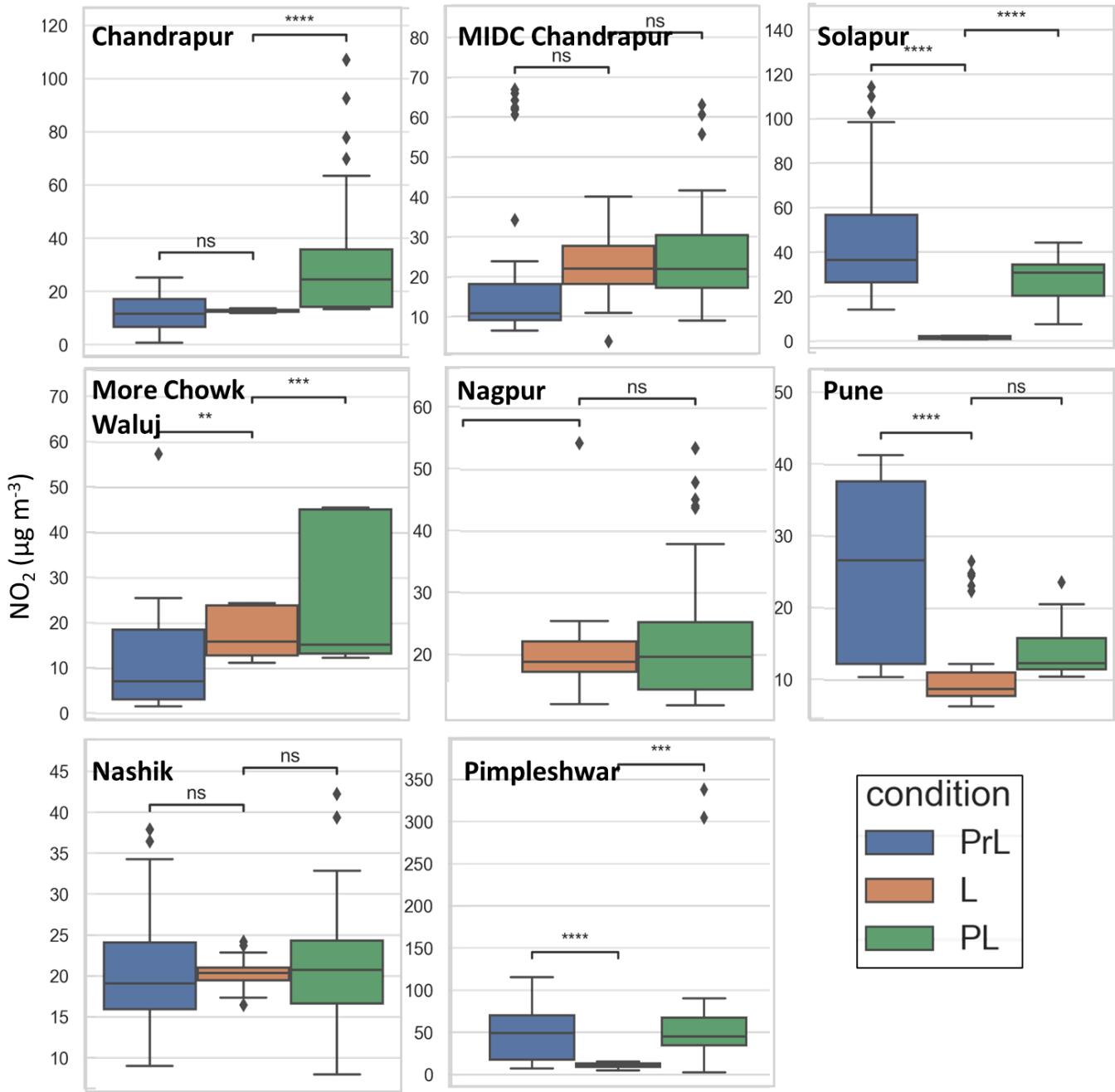


Figure 4.12: Comparison NO_2 during lockdown month (April). Pre lockdown years (PrL) 2017 - 2019, Lockdown – 2020, Post Lockdown – 2021 -2022. (Concentration of NO_2 is in $\mu\text{g}/\text{m}^3$).
Annotations: ns: $p \leq 1$; *: $1e-2 < p \leq 5e-2$; **: $1e-3 < p \leq 1e-2$; ***: $1e-4 < p \leq 1e-3$; ****: $p \leq 1e-4$.

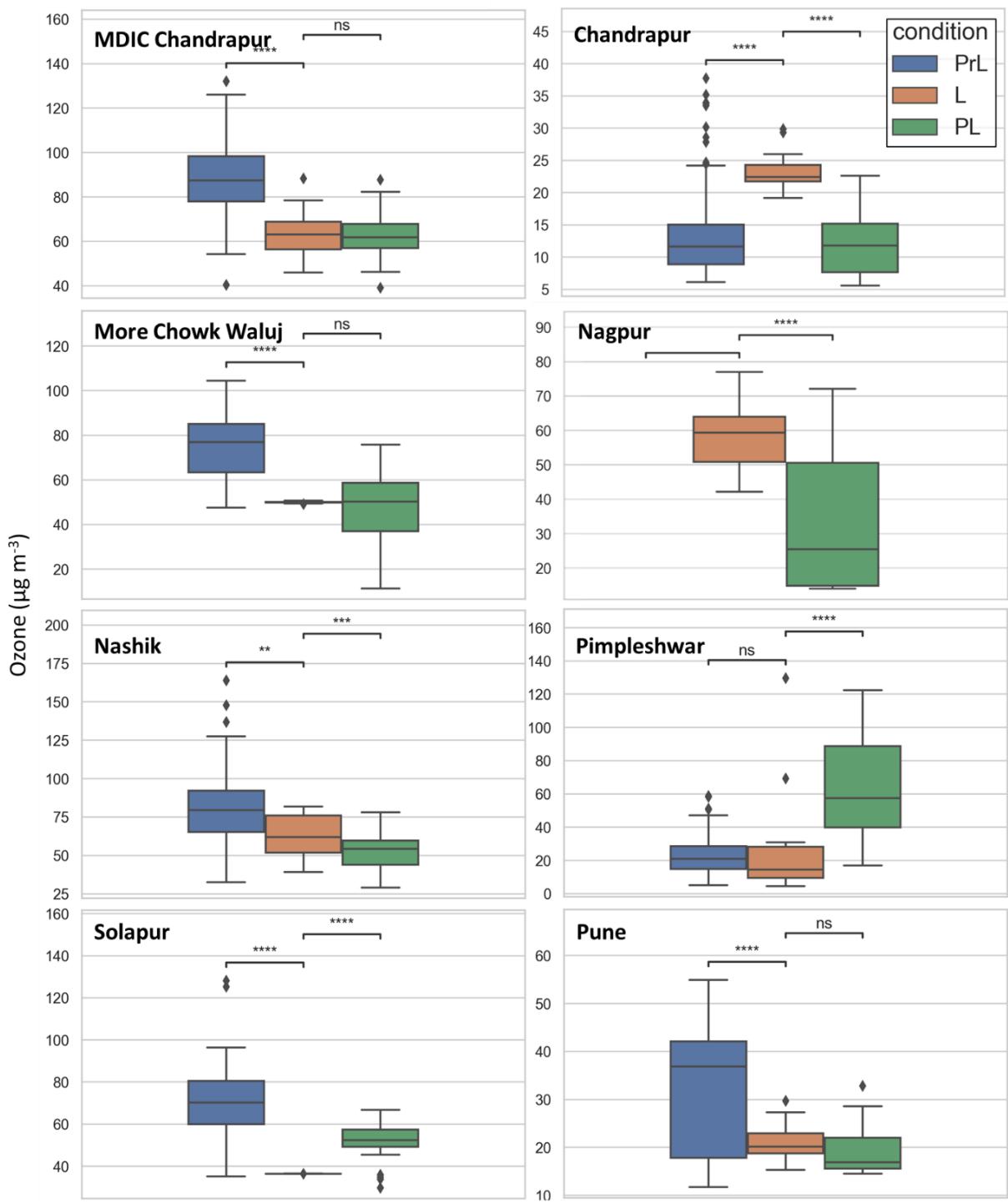


Figure 4.13: Comparison O₃ during lockdown month (April). Pre lockdown years (PrL) 2017 - 2019, Lockdown – 2020, Post Lockdown – 2021 -2022. (Concentration of NO₂ is in $\mu\text{g/m}^3$).

Annotations: ns: p <= 1; *: 1e-2 < p <= 5e-2; **: 1e-3 < p <= 1e-2; ***: 1e-4 < p <= 1e-3;

****: p <= 1e-4.

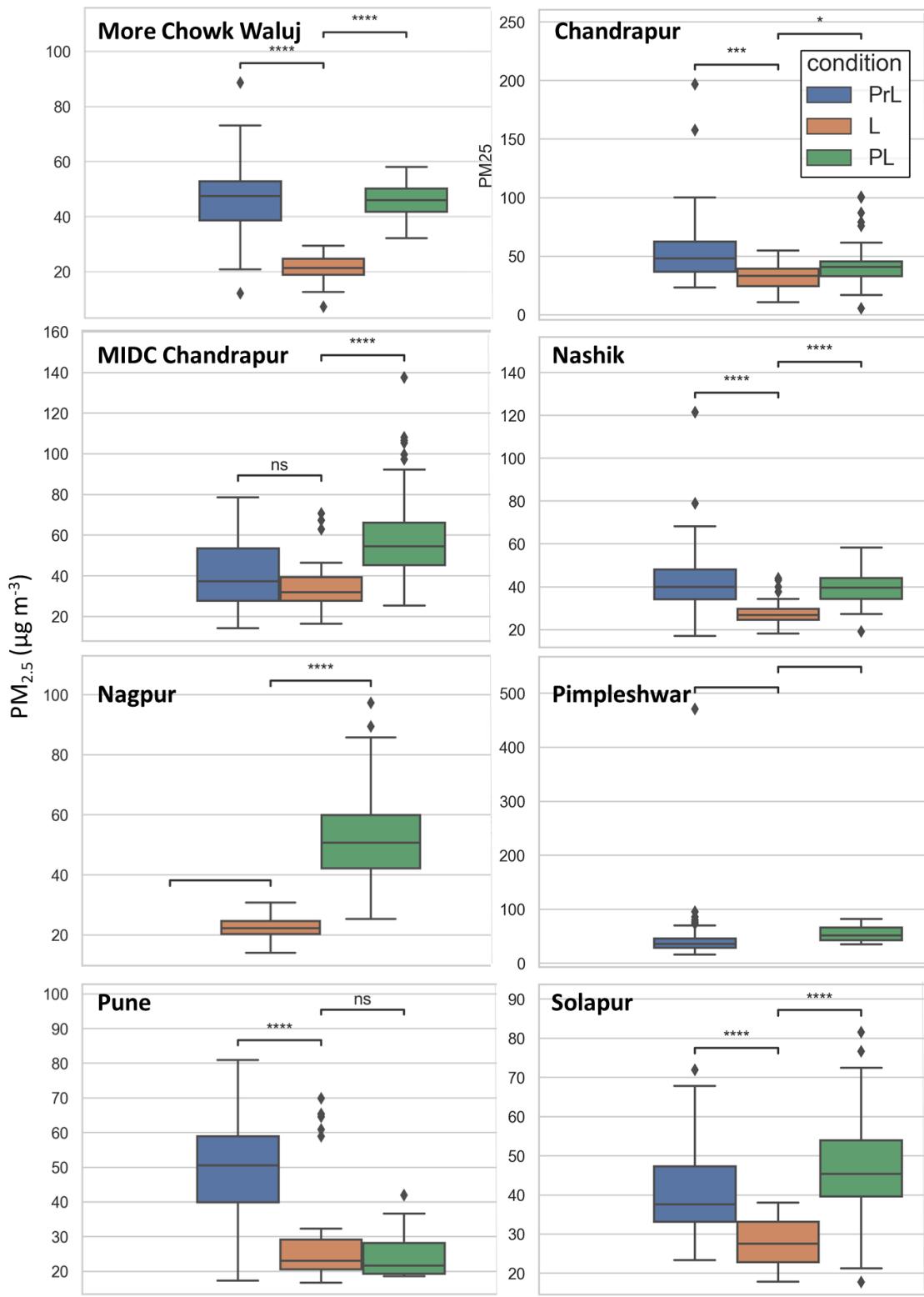


Figure 4.14: Comparison PM₁₀ during lockdown month (April). Pre lockdown years (PrL) 2017

- 2019, Lockdown – 2020, Post Lockdown – 2021 -2022. (Concentration of PM₁₀ is in $\mu\text{g}/\text{m}^3$).

Annotations: ns: $p \leq 1$; *: $1\text{e-}2 < p \leq 5\text{e-}2$; **: $1\text{e-}3 < p \leq 1\text{e-}2$; ***: $1\text{e-}4 < p \leq 1\text{e-}3$;

****: $p \leq 1\text{e-}4$.

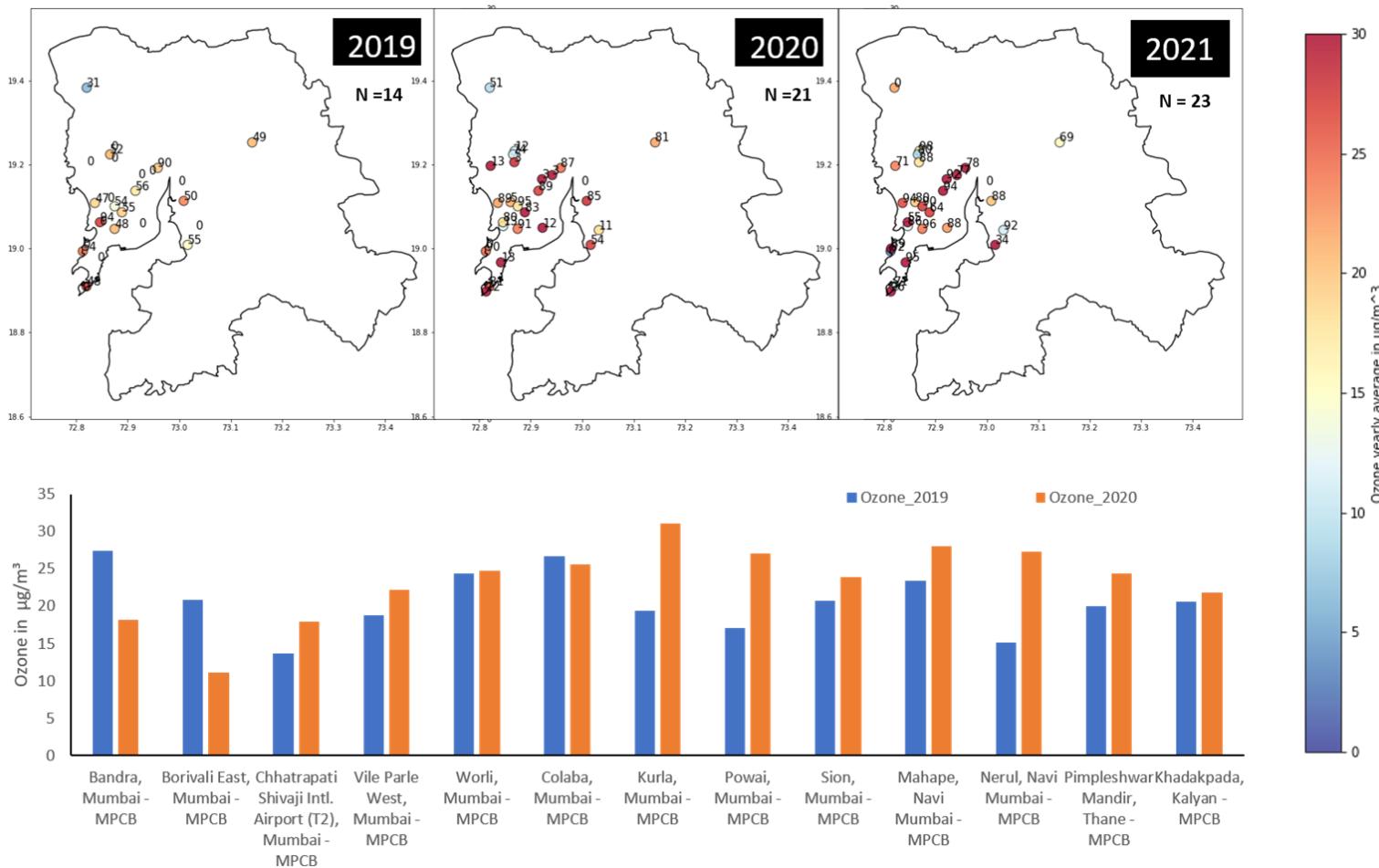


Figure 4.15: Comparison of O_3 from 2019 to 2021. Pre lockdown year - 2019, Lockdown – 2020 (Concentration of O_3 is in $\mu\text{g}/\text{m}^3$). Annotations on the marker denote the data availability used to calculate the yearly mean. N denotes the total number of stations with data available.

Chapter 5

Validation of satellite-based ground level PM_{2.5} estimation

5.1 Methodology:

5.1.1 Hybrid PM_{2.5} estimates

The PM_{2.5} estimates utilized in this study were derived from the work of Van Donkelaar et al. (2021), which primarily employed the physical relationship (η) between Aerosol Optical Depth (AOD) and ground-based measurements of PM_{2.5}. Multiple satellite products, including the Multi-angle Imaging Spectroradiometer (MISR), Moderate Resolution Imaging Spectroradiometer (MODIS) Dark Target, MODIS and Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Deep Blue, and MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC) with simulated AOD from the GEOS-Chem chemical transport model (<http://geos-chem.org>; V11-01), were used to retrieve AOD level 2 data (Van Donkelaar et al., 2021). The daily AOD values at a $0.1^\circ \times 0.1^\circ$ grid resolution were averaged within predefined grids on a monthly basis (AOD_{monthly,k}). Unavailable grid values were replaced by the averages of available AOD values after scaling, taking into account their monthly variability. Additionally, the AOD_{monthly,k} values were weighted (w_k) based on their agreement with AOD measurements from Aerosol Robotic Network (AERONET) observations, resulting in the AOD_{monthly,best,k} values. The weighting was calculated using the following equation:

$$w_k = \text{NRMSD}_k^{-1} \times AOD_k / (AOD_{adj,k} - AOD_k)$$

Here, NRMSD_k represents the normalized root mean square difference, AOD_{adj,k} is the AOD_k value adjusted through regression in comparison with AERONET AOD. The AOD_{monthly,best,k} values and the simulated AOD_{monthly,best,k} were then averaged to obtain an overall best estimate.

The simulated AOD at the monthly level was derived by establishing a geophysical relationship between AOD and ground-based measurements of PM_{2.5}, where η represents a spatially and temporally varying ratio between AOD and the monthly average of ground-based PM_{2.5} concentrations. The GEOS-Chem calculation also incorporates various factors such as sampling time, composition, emissions, hygroscopicity, and meteorological variables, such as relative humidity, wind speed, boundary layer height and wind direction.

5.1.2. Validation between PM_{2.5} satellite estimates and ground-based records

The evaluation of satellite-derived PM_{2.5} estimates (PM_{2.5, sat}) was conducted by comparing them with ground-based PM_{2.5} measurements (PM_{2.5, GM}) at a monthly resolution. The PM_{2.5, GM} data were pre-processed at a 15-minute resolution using the AirPy python tool (Chapter 3) (<https://github.com/Madhumitha11-s/AirPy>) for 1 site in Kalyan, 6 in Mumbai city, 14 in Mumbai suburban, 5 in Thane for all available PM_{2.5, GM} data between 2019-2021. The analysis revealed a strong relationship between PM_{2.5, sat} and PM_{2.5, GM}, as indicated by a reduced major axis (RMA) slope ranging from 0.75 to 1.0 and a high correlation coefficient of 0.96 to 0.98 (Figure 5.1).

5.1.3. Model performance before recalibration of Satellite measurements

Despite the favourable slope and correlation coefficient, the coefficient of determination (R-squared) between PM_{2.5, GM} and PM_{2.5, sat} was estimated to be 0.39. This relatively low value suggests that only 39% of the variability in the ground-based measurements can be explained by satellite estimates. The poor agreement observed between the two datasets could be attributed to the calibration procedure employed in this study. The calibration was conducted using a geographically weighted regression (GWR) technique, which utilized global ground-based observations from monitoring sites predominantly located in Canada, the United States, and China. The limited representation of monitoring sites across the globe might have led to

discrepancies between the satellite estimates and ground-based measurements, resulting in the observed low coefficient of determination.

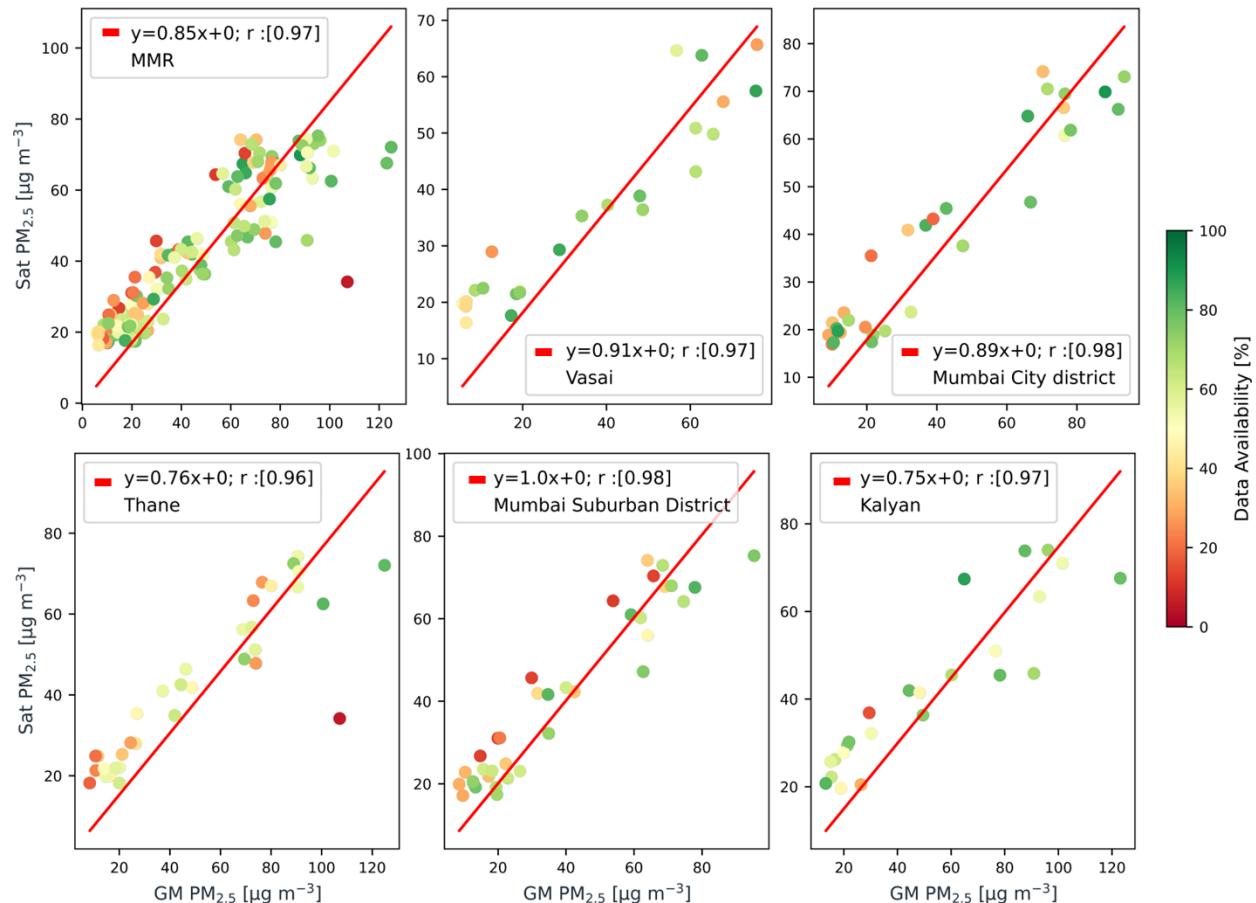


Figure 5.1: Comparison of PM_{2.5}, sat and PM_{2.5}, GM from 2019-2021. Each data point in the scatterplot corresponds to monthly averages at each individual district. The PM_{2.5}, GM of each district was average of one or more station functional between 2019-2021. The color of each scatter point denotes the amount of data available at the time of calculation of monthly average at each CPCB site which is again aggregated at district levels.

To mitigate the limitations arising from the calibration procedure, an alternative approach was adopted. We decided to weigh the PM_{2.5}, sat estimates using the corresponding PM_{2.5}, GM values. By incorporating the PM_{2.5}, GM data as a weighting factor, we aimed to refine the satellite estimates and enhance their accuracy. This weighting approach accounted for the agreement

between the satellite estimates and the ground-based measurements, offering a means to incorporate the valuable information provided by the local ground-based observations in MMR into the satellite-derived estimates.

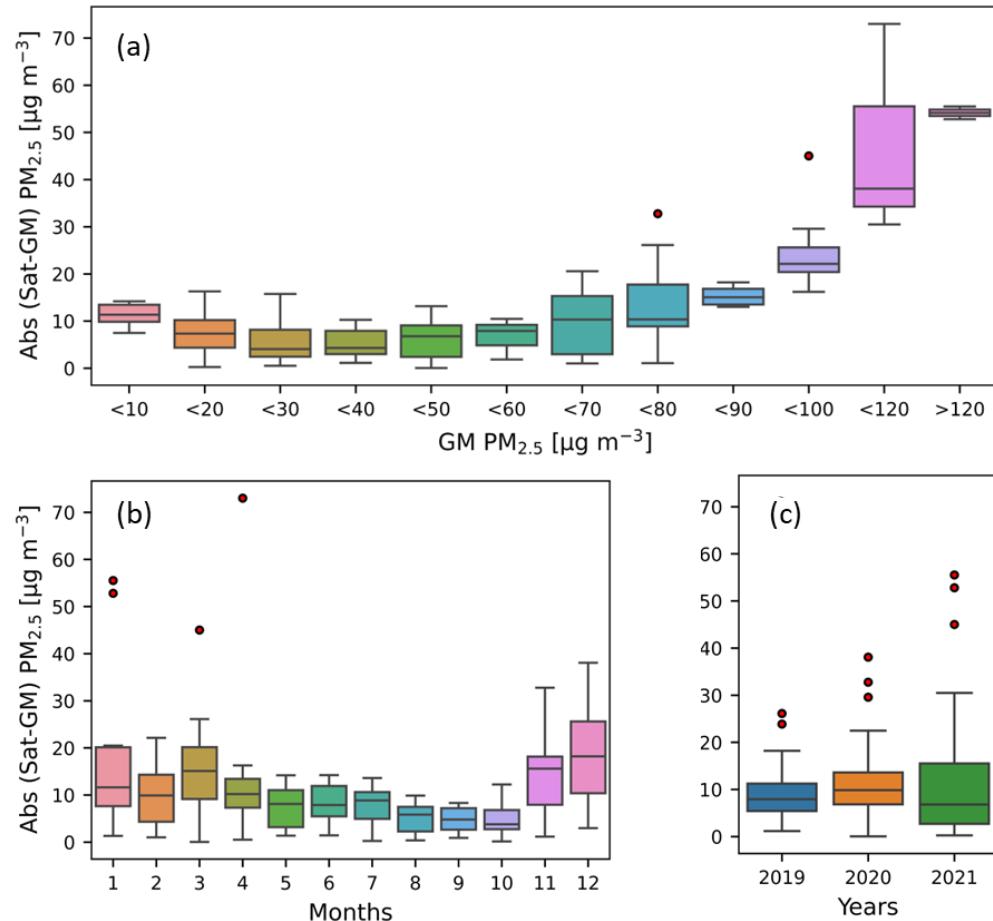


Figure 5.2: Analysis of PM_{2.5}, sat Discrepancy: Ground-Based Measurements vs. Satellite Estimates (2019-2021). Panel (a) illustrates the dependency of error bias on the concentration of PM_{2.5, GM}. Panel (b) explores the independence of the bias on monthly variability, while panel (c) investigates its independence on annual variability.

In addition to evaluating the agreement between the datasets, it was important to analyze the behaviour of the error in the retrieved PM_{2.5}, sat dataset across a wide range of PM_{2.5, GM} concentrations. To this end, a retrieval bias was calculated as the difference between the PM_{2.5}, GM values obtained from the CPCB sites and the corresponding PM_{2.5}, sat estimates. This bias

was then plotted as a function of the PM_{2.5, GM} concentrations, yielding valuable insights into the performance of the satellite-derived estimates. The analysis of the retrieval bias as a function of PM_{2.5, GM} concentration (as depicted in Figure 5.2) revealed interesting patterns.

5.1.4. Concentration dependent bias in Satellite estimates

For PM_{2.5} levels up to 80 μgm^{-3} , the median bias remained consistently below 10 $\mu\text{g m}^{-3}$ (Figure 5.2a). This indicates that, on average, the satellite estimates exhibited a slight underestimation compared to the ground-based measurements but within an acceptable range similar to studies by Dey et al. (2020). However, beyond the threshold of 80 $\mu\text{g m}^{-3}$, the underestimation in PM_{2.5, sat} became more pronounced, suggesting that the satellite estimates progressively deviated from the ground-based measurements at elevated PM_{2.5} concentrations. Further investigation into the discrepancy between PM_{2.5, sat} and PM_{2.5, GM} was conducted by examining the variation of this discrepancy with respect to different factors. We observed that the difference between the satellite estimates and ground-based measurements remained relatively consistent throughout the year, regardless of seasonal or temporal variations (Figure 5.2b-c).

To address the issue of increasing discrepancy at higher PM_{2.5, sat} values, a basic random forest (RF) model was developed to train the PM_{2.5, sat} using only the PM_{2.5, GM} values, enabling the prediction of ground-based measurements referred to as PM_{2.5, adj sat}. Specifically, the RF model was trained using the PM_{2.5, GM} data from Vasai and Thane, covering the period from 2019 to 2021. Only two districts were chosen to ensure the model's generalizability.

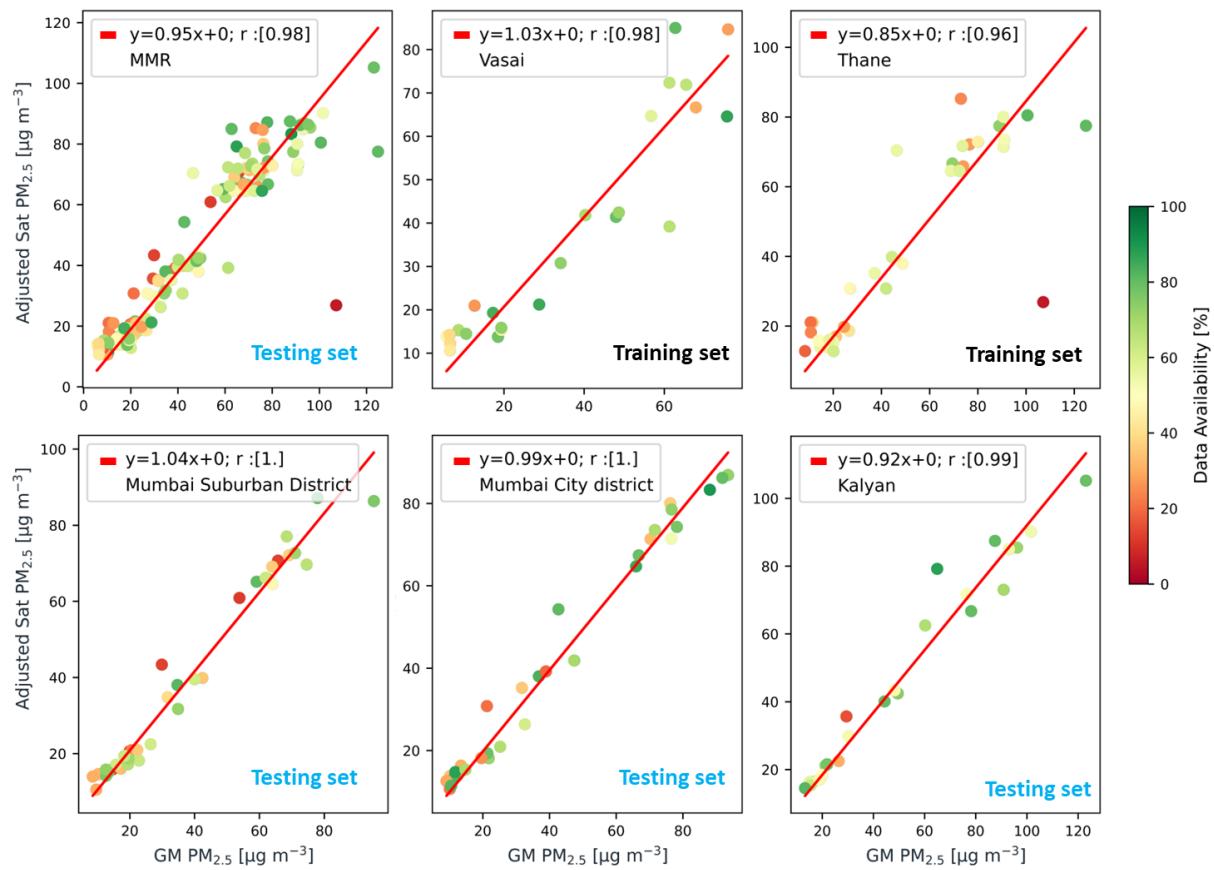


Figure 5.4: Comparison of PM_{2.5}, Adj sat and PM_{2.5}, GM from 2019-202. This figure illustrates the comparison between PM_{2.5} estimates obtained from the adjusted satellite data (PM_{2.5}, Adj sat) and ground-based measurements (PM_{2.5}, GM) for the period spanning 2019 to 2021. Each data point in the scatterplot represents the monthly averages for individual districts. The PM_{2.5}, GM values for each district were calculated as the average of one or more functional monitoring stations operating between 2019 and 2021. The color coding of each scatter point indicates the availability of data during the calculation of monthly averages at each CPCB site, which were then aggregated at the district level. Notably, only the PM_{2.5} estimates from the Vasai and Thane districts were used to train the model for predicting PM_{2.5}, GM values for all three years. Furthermore, predictions for the years 1994 to 2020 were also generated based on this trained model.

5.1.3. Model performance after recalibration of Satellite measurements

The implementation of the RF model yielded notable improvements in the relationship between $\text{PM}_{2.5, \text{adj sat}}$ and $\text{PM}_{2.5, \text{GM}}$ (Figure 5.4). Both the training and testing datasets exhibited enhanced agreement, as evidenced by a significantly improved slope in the relationship. Moreover, the correlation analysis between $\text{PM}_{2.5, \text{adj sat}}$ and $\text{PM}_{2.5, \text{GM}}$ demonstrated a strong positive correlation, characterized by a high coefficient of determination (R^2) of 0.84 and a coefficient of correlation (r) of 0.93. These statistical metrics indicate that the adjusted satellite estimates obtained through the RF model were in strong agreement with the ground-based measurements, further validating the effectiveness of this modelling approach. Notably, the post-adjustment of $\text{PM}_{2.5, \text{sat}}$ using the RF model led to a significant reduction ($p\text{-value} < 0.05$) in the error bias, even for concentrations exceeding $80 \mu\text{g/m}^3$.

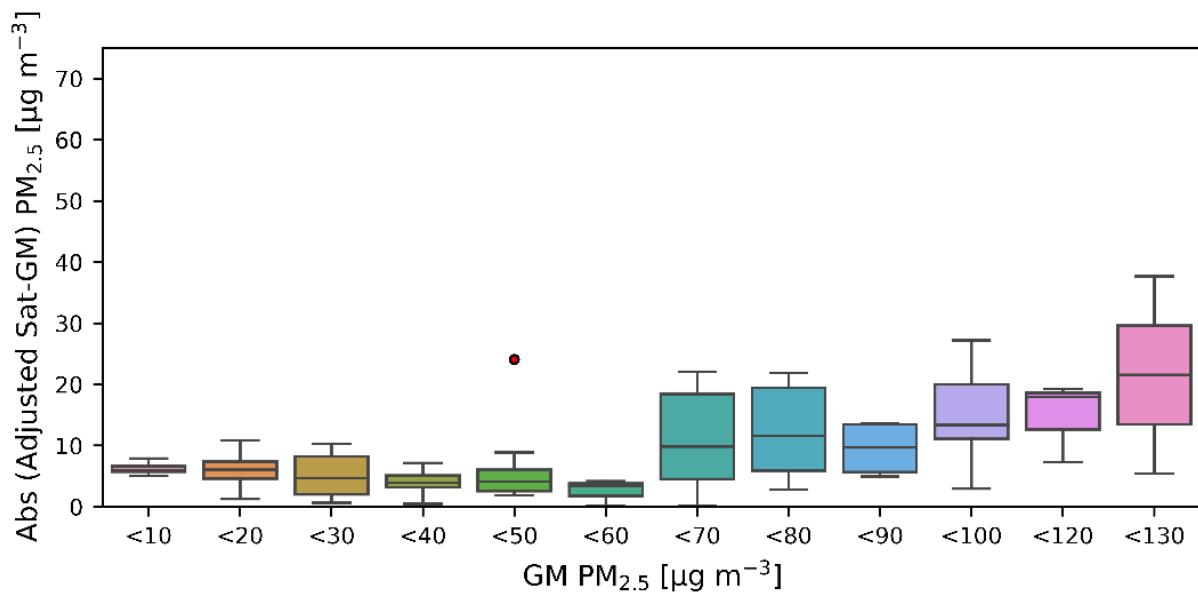


Figure 5.3: Improvement in agreement of $\text{PM}_{2.5, \text{sat}}$ with $\text{PM}_{2.5, \text{GM}}$ post adjustment: Ground-Based Measurements vs. Satellite Estimates (2019-2021). Panel a illustrates the dependency of error bias on the concentration of $\text{PM}_{2.5, \text{GM}}$. Panel b explores the independence of the bias on monthly variability, while panel c investigates its independence on annual variability.

Chapter 6

The Impact of COVID-19 and Post-COVID Trends on Air Quality in the MMR

6.1 Predicting Business-as-Usual PM_{2.5} Levels for COVID Years

6.1.1 Business-as-Usual (BAU) Model development for COVID Years

In this study, we employed a RF model to predict BAU PM_{2.5, GM} levels during the COVID years, specifically for lockdown years 2020 and 2021. The RF model was trained using historical PM_{2.5} adjusted satellite PM_{2.5} measurements (PM_{2.5, adj sat}) from 1994 to 2018 at a monthly level. By harnessing the power of ensemble learning, we utilized 1000 decision trees with a minimum number of leaf nodes as 10 in the RF model and implemented a bagging technique. To ensure robustness, we constructed 100 different RF models individually for each site and for each monthly average prediction. This approach is similar to studies from Wang et al., (2020) and involves varying the sample set with repetition and random state.

We used weather data from the ERA5 reanalysis dataset ²⁵ as predictor variables for this RF model. These variables encompassed a range of meteorological parameters, including boundary layer height, total cloud cover, surface net solar radiation, total precipitation, temperature, relative humidity, wind speed, wind direction, and atmospheric pressure, as well as time variables represented as integers (month and year). By incorporating these factors as inputs, we aimed to capture the intricate interactions between meteorological conditions and PM_{2.5} concentrations.

To validate the efficacy of our developed RF model in predicting Business-as-Usual (BAU) conditions, we excluded the 2019 data from the training set and tested the model's performance on this latest pre-COVID year (2019). By comparing the RF model predicted PM_{2.5, adj sa} with

the observed PM_{2.5}, adj sat we obtained an impressive coefficient of determination (R-squared, R²) of 0.88, indicating a strong correlation between the predicted and observed PM_{2.5} levels.

Furthermore, the Normalized Root Mean Square Error (NRMSE) was found to be exceptionally low, nearing 0.09. These findings underscore the high accuracy and reliability of our RF model in capturing the complex dynamics of PM_{2.5} concentrations during BAU periods.

Building on the successful performance of the RF model in predicting BAU for 2019, we extended its application to forecast PM_{2.5} levels during the subsequent years, 2020 and 2021. These years were marked by the imposition of stringent lockdown measures in India due to the COVID-19 pandemic. By employing the same RF model, we aimed to capture the expected PM_{2.5} levels under these unique circumstances.

6.1.2 Feature importance in Business-as-Usual RF model

In our study, we conducted feature importance analysis to assess the contribution of each variable in predicting air pollution levels (Figure 6.1). Feature importance is a techniques that assign scores to input features in a predictive model, thereby indicating the relative significance of each feature in making predictions (scikit-learn 1.2.2: https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html). This analysis helps determine the relevance of specific variables for the model and its predictions. In a machine learning model, the prediction of the target variable is influenced by multiple input variables. The goal of Partial dependence plots (PDPs) is to isolate the effect of one particular variable and visualize its impact on the predicted outcome. Using PDP, among all the weather components, we found that wind direction, u wind (eastward wind), and atmospheric pressure are significant factors in the model (Figure 6.2a). Specifically, negative u wind indicates the presence of eastward winds, while negative v wind suggests northward winds (Figure 6.2b), both of which have been associated with an increase in pollution levels. Additionally, we observed that heatwaves often coincide with high atmospheric pressure, leading to the

formation of a stagnant layer of air near the ground. In urban areas, this stagnant layer can trap air pollutants, resulting in higher concentrations of pollutants in the air.

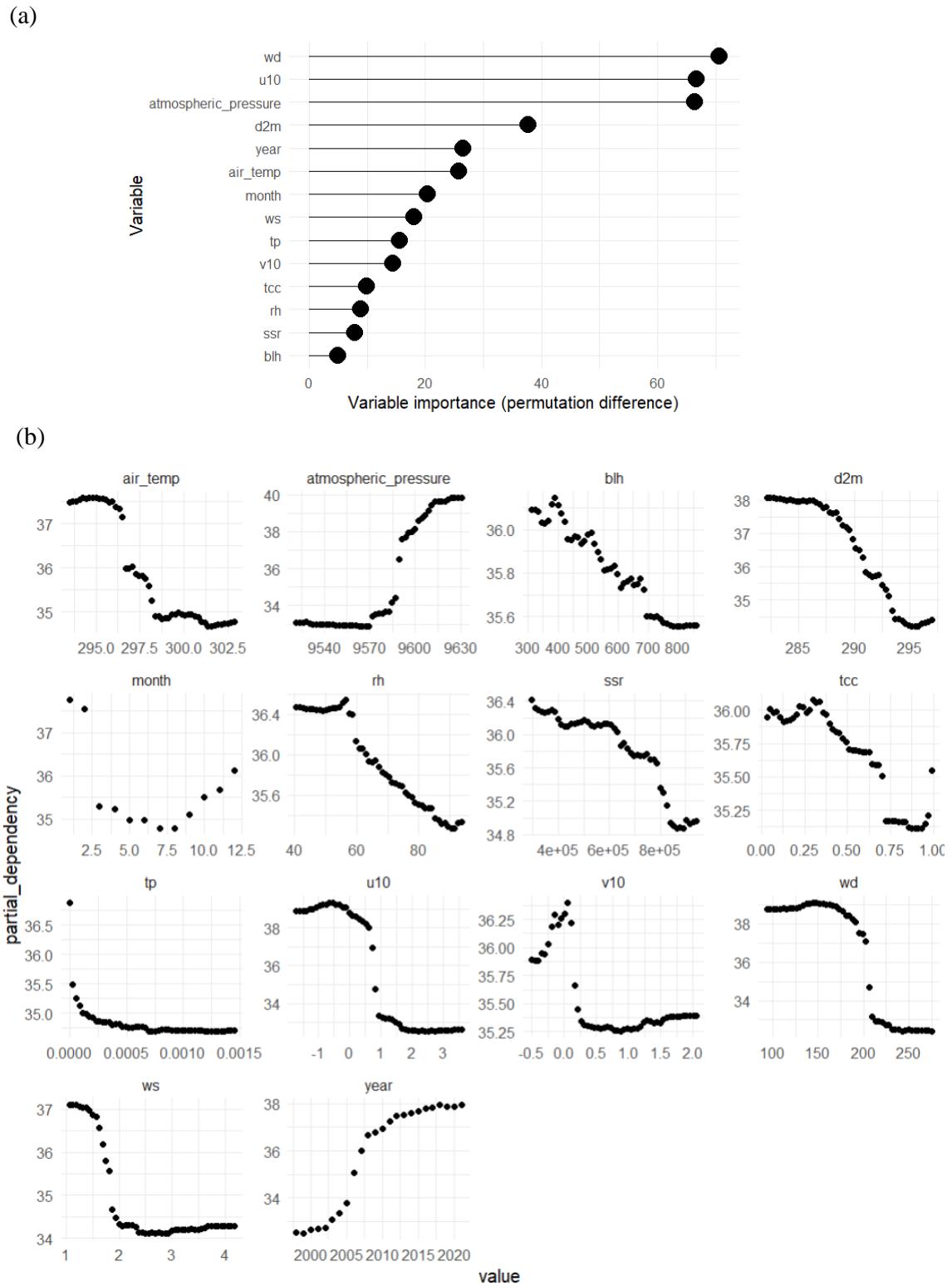


Figure 6.1. Variable importance and partial dependency plot for RF BAU model.

Panel (a) shows the importance of features in the model, whereas panel (b) shows the effect of each feature in the model's performance.

6.2 Weather and PM_{2.5} trend over past two decades.

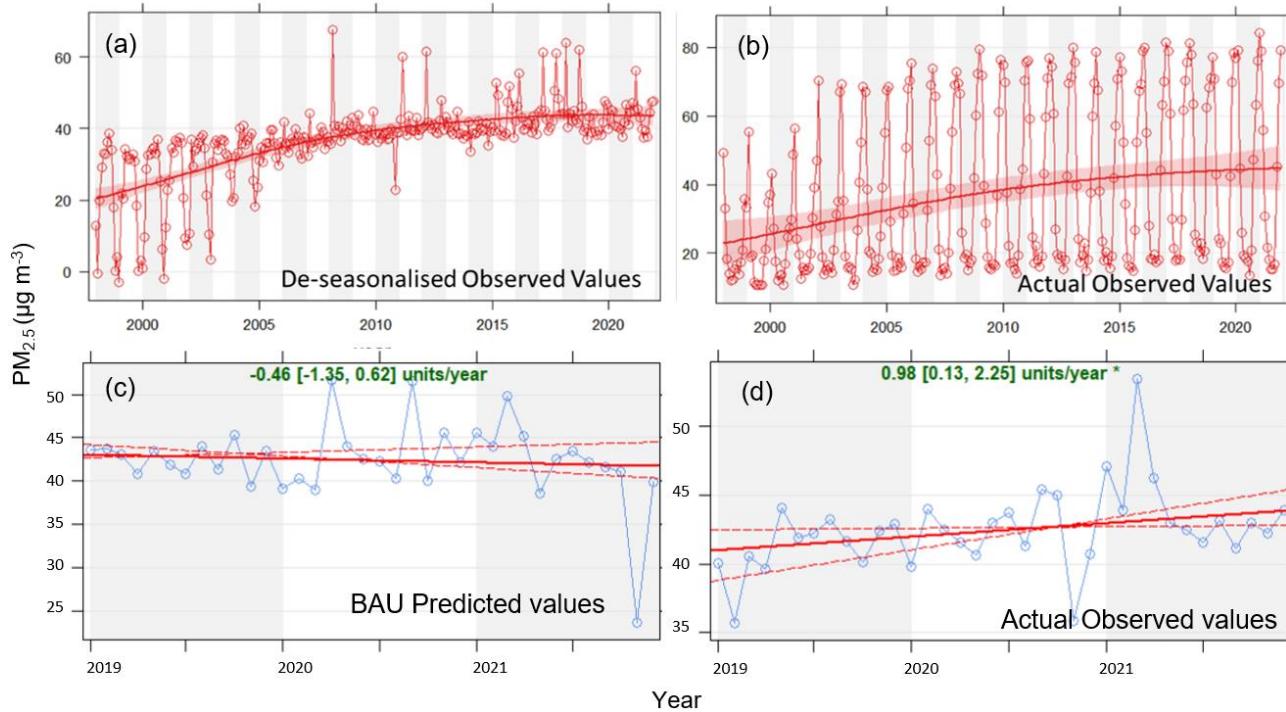


Figure 6.2. Trends in PM_{2.5} between 1994-2021 using the after smoothening the pollutant data from MMR region. The shading in (a) & (b) shows the estimated 95 % confidence intervals. The solid red line shows Theil Sen, and the dashed red lines show the 95 % confidence intervals for the trend based on resampling methods for 2019-2021 in observed PM_{2.5} in (d) and predicted PM_{2.5} in (c). * Indicates the significance of the trend.

To analyze and understand the trend of PM_{2.5} pollution in the Mumbai Metropolitan Region more comprehensively, we employed the Seasonal-Trend decomposition using LOESS (STL) method using OpenAir package in R programming (<https://github.com/davidcarslaw/openair>). This statistical technique effectively separates a time series into three distinct components: trend, seasonality, and residual.

The trend component captures the long-term pattern or underlying direction of the time series. In the context of PM_{2.5} pollution, it helps identify the overall increasing or decreasing trend in pollution levels over time. Through the application of STL, we were able to observe that the PM_{2.5} levels exhibited a substantial increase at a higher speed from the early 2000s until 2009, reflecting the period of rapid urbanization and industrial growth (Figure 6.2). However, starting from 2009, the trend component indicated a stabilization of PM_{2.5} levels, with the rate of increase in pollution becoming significantly less pronounced. Brihanmumbai Municipal Corporation (BMC) also suggested that PM_{2.5} pollution in Mumbai has reduced a lot and showed a downward trend between 2015-2021 ¹⁰⁰, which contradicts our results where we see a significant upward trend in PM_{2.5} with rate of 0.98 µgm⁻³/year (Figure 6.2d).

The utilization of the TheilSen function in our study provided a robust framework for analyzing trends in air pollution data, enabling a comprehensive investigation of PM_{2.5} concentration patterns within distinct wind sectors. This approach allowed us to delve into the underlying dynamics and elucidate the factors contributing to significant upward or downward trends in pollutant levels.

Our analysis revealed a substantial and rapid increase in PM_{2.5} concentrations from the early 2000s to 2009. This notable rise suggests the presence of intensified pollution sources and activities during that period. However, a notable shift occurred in 2009, characterized by the attainment of a concentration plateau and a subsequent deceleration in the rate of pollution increase. This change could be attributed to a combination of factors, including the implementation of pollution control measures and/or population stabilization.

Further examination of the sources driving this observed increase in pollution indicated a significant contribution from pollutants transported by winds originating from the east, south, and southeast directions. Among these, the south-east wind emerged as the primary contributor,

with an annual increase in PM_{2.5} concentrations of 1.77 µg/m³ (Figure 6.3). Additionally, the south and east winds exhibited substantial impacts, contributing to annual increases of 1.34 µg/m³ and 1.17 µg/m³, respectively. In contrast, the west and southwest winds made comparatively minor contributions, with annual increases in PM_{2.5} concentrations of less than 0.55 µg/m³. Moreover, our analysis highlighted the noteworthy influence of the southwest wind direction in reducing pollution levels. Figure 6.4 visually represents this finding, demonstrating a statistically significant decrease in pollution over the past five years (*p*-value < 0.001) associated with the southwest wind. Specifically, this wind direction contributed to an annual mean of PM_{2.5} concentrations by 1.05 µg/m³ per year. These findings underscore the critical role of wind patterns in modulating pollutant dispersion and accumulation, thereby exerting a tangible influence on regional air quality.

The Conditional Probability Functions (CPF) approach was initially employed to identify the prevailing wind directions associated with high concentrations of a specific pollutant ¹⁰¹. It provides valuable insights into the likelihood of observing such elevated pollutant levels based on different wind directions. In our study, we utilized CPF to determine the specific intervals of wind direction and wind speed that are predominantly associated with high pollutant concentrations and quantified the probability of their occurrence. However, it is crucial to note that the presence of potential pollution sources within the percentile ranges of 0 to 50 cannot be dismissed as mere artifacts ^{102,103}. These sources may originate from more distant point sources whose plumes have undergone greater dilution over extended distances. In conventional polar plots, such sources might be obscured or overshadowed. Consequently, it is plausible to consider that these sources contribute significantly to pollutant levels within the mentioned percentile ranges. Furthermore, road sources could potentially play a crucial role in pollutant concentrations falling between the 60th and 75th percentiles.

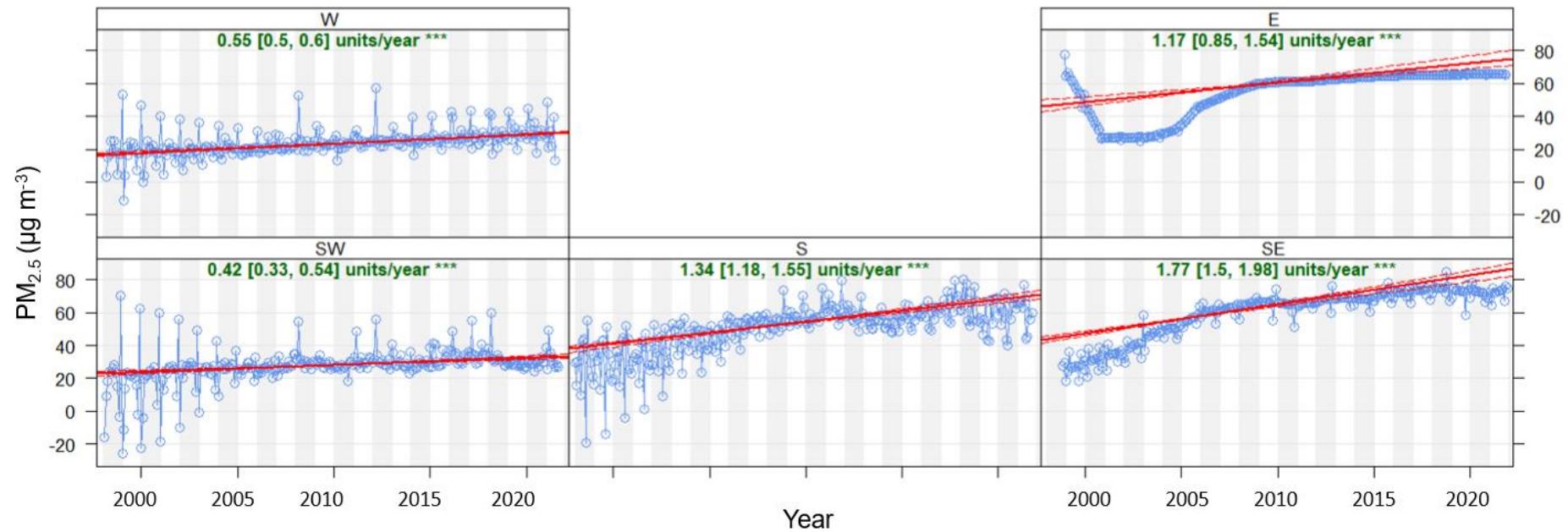


Figure 6.3. Trends in PM_{2.5} at MMR region from 1994-2021. The plot shows the deseasonalised monthly mean concentrations of PM_{2.5}. The solid red line shows the trend estimate by TheilSen and the dashed red lines show the 95 % confidence intervals for the trend based on resampling methods. For example, the overall trend from Southeast direction is shown at the top-left as 1.77 (μgm^{-3}) per year and the 95 % confidence intervals in the slope from 1.5–1.98 μgm^{-3} /year. The *** show that the trend is significant to the 0.001 level.

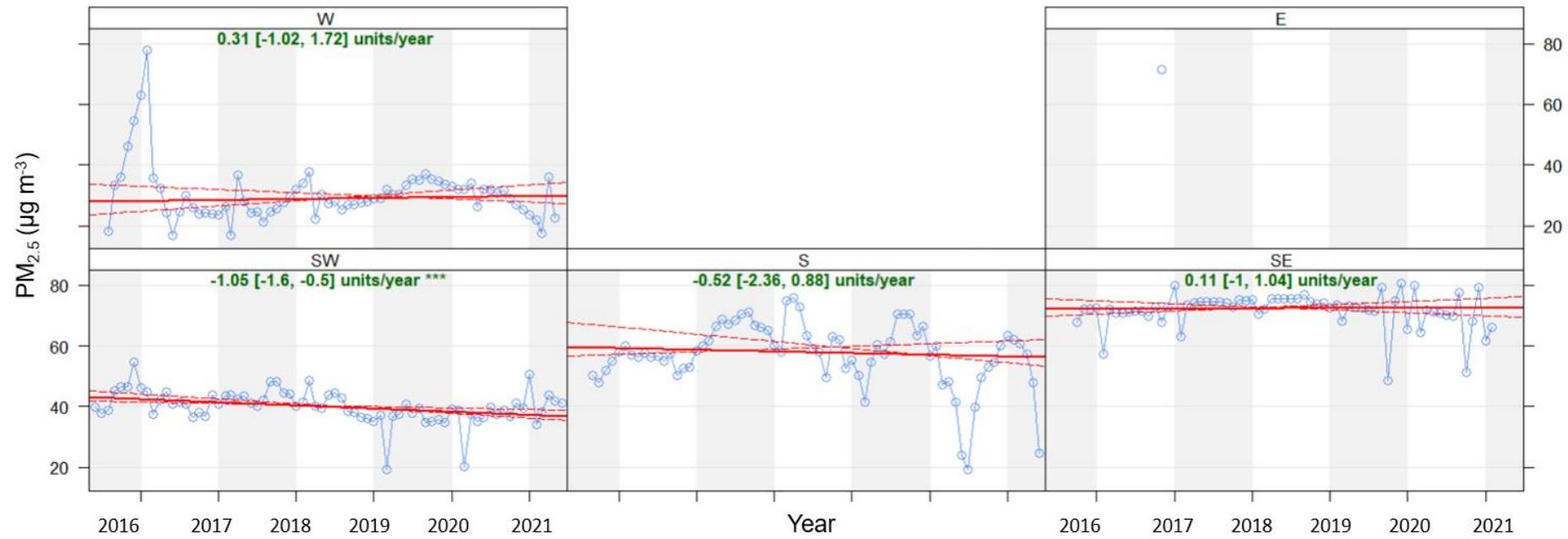


Figure 6.4. Trends in PM_{2.5} at MMR region from 2016-2021. The plot shows the deseasonalised monthly mean concentrations of PM_{2.5}. The solid red line shows the trend estimate by TheilSen and the dashed red lines show the 95 % confidence intervals for the trend based on resampling methods. For example, the overall trend from the Southeast direction is shown at the top-left as 1.77 μgm^{-3} per year and the 95 % confidence intervals in the slope from 1.5–1.98 $\mu\text{gm}^{-3}/\text{year}$. The * * * show that the trend is significant to the 0.001 level.

The chimney stacks and other stationary emission sources become particularly influential in contributing to higher pollutant percentiles, ranging from the 75th to the 100th percentiles. This highlights the significance of industrial emissions and point sources in driving the observed high concentrations of pollutants within these percentile ranges^{102,103}. Figure 6.5 reveals changes in the hotspot below the center of the MMR for the 75th to 100th percentile range in the CPF plot. These changes coincided with the COVID-19 pandemic, resulting in reduced industrial activity and decreased vehicular emissions. During this period, a noticeable reduction in the hotspot within the 75th to 100th percentile range is observed, attributed to decreased industrial activities. The temporary closures and limited operations of industries led to decreased emissions, resulting in lower pollutant concentrations within this percentile range. Similarly, the CPF plot also shows a decrease in the 50th to 75th percentile range, which can be attributed to reduced vehicular emissions due to restricted mobility and decreased traffic volume during the pandemic.

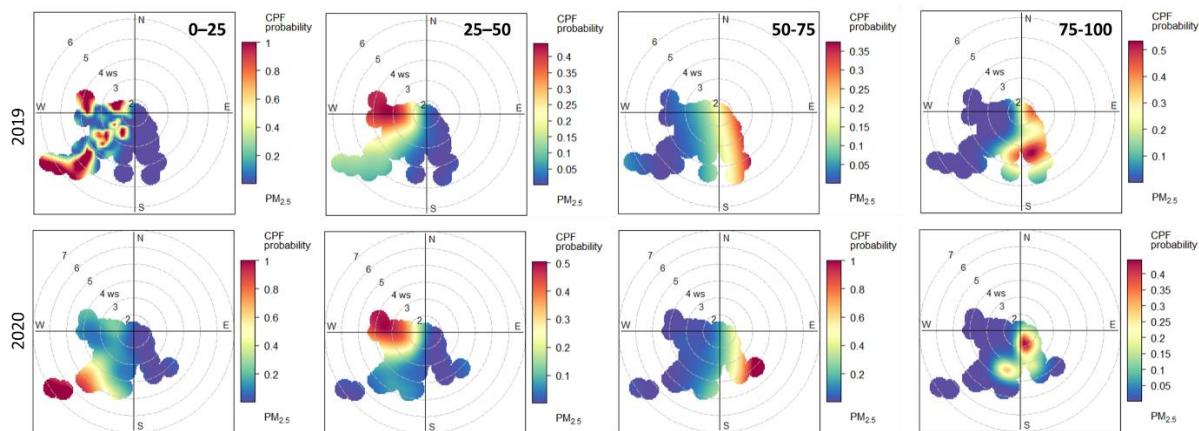


Figure 6.5. PolarPlot of PM_{2.5} concentrations at MMR based on the CPF function for a range of percentile intervals from 0–25, 25–50, 50–75, 75–100 (left to right) for 2019 and 2020.

The relationship between wind speed and pollutant concentrations can be intricate and multifaceted. It is not always the case that higher wind speeds lead to lower concentrations due to the presence of recirculation phenomena¹⁰⁴. Bivariate polar plots prove to be effective tools

in unraveling these complexities and providing insights into the spatial distribution of pollutant concentrations. In Figure 6.6, the polar plot clearly demonstrates that the highest PM_{2.5} concentrations were observed when the wind originated from the south-west and south directions in 2019. These wind patterns appeared to contribute significantly to the elevated pollution levels during that period. However, notable changes were observed during the COVID-19 year of 2020. Specifically, the pollution originating from the south direction exhibited a minimal impact compared to previous years. This reduction in pollution levels from the south can be attributed to the decreased industrial activity and reduced vehicular emissions resulting from the pandemic-induced restrictions. Furthermore, the observed pollution levels originating from the south-east direction were lower than what was predicted by the Business as Usual (BAU) model. This finding suggests that factors other than the conventional sources considered in the BAU model may have influenced the pollution levels from the south-east direction. These additional factors could include localized sources or changes in emission patterns that were not adequately captured by the model. The average windspeed between 2016-2021 was 2.68 ms⁻¹ and always less than 2.78 except for the year 2019 where windspeed was 2.81 ms⁻¹, which could have contributed to reduced pollution during 2019 (Figure 6.7).

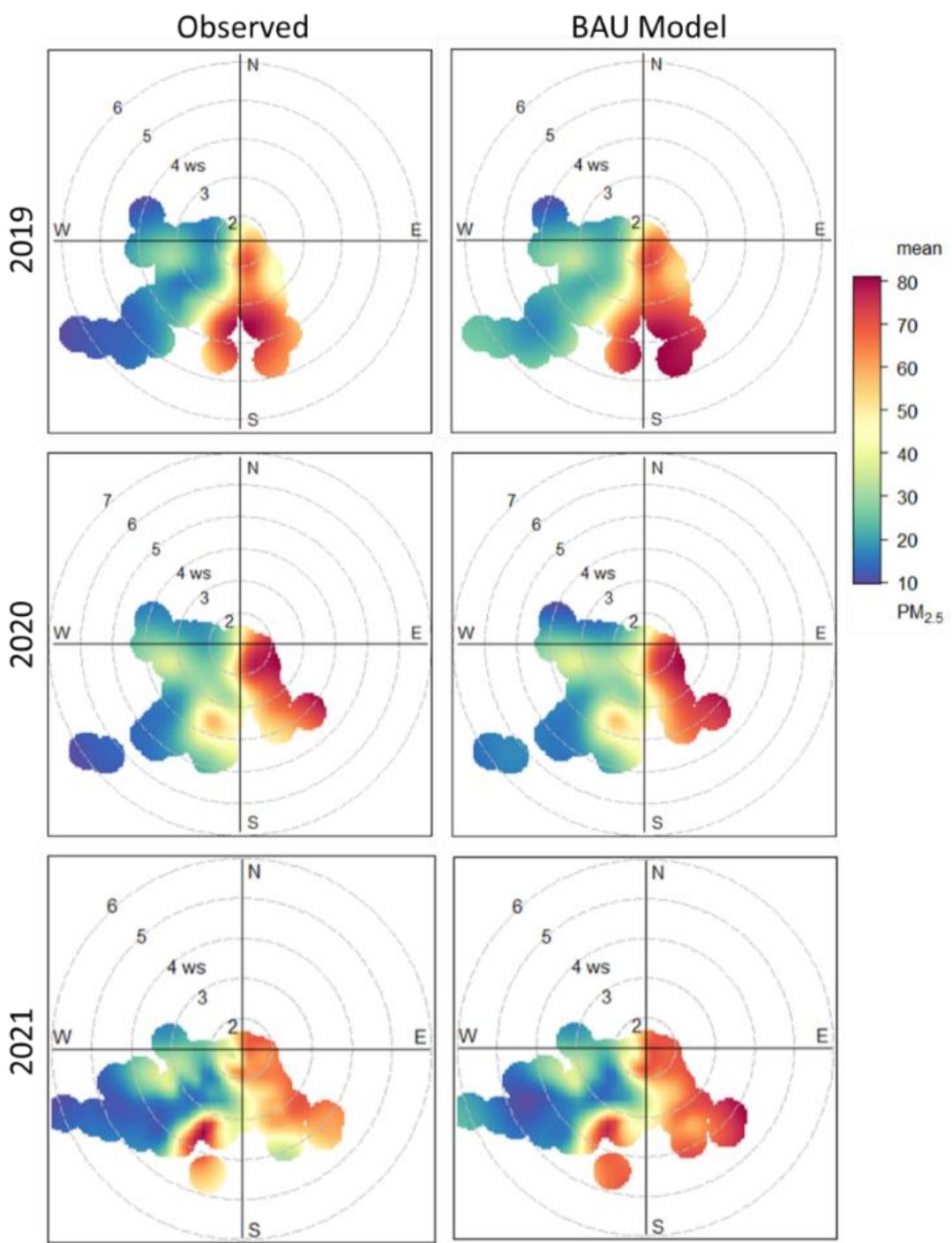


Figure 6.6: PolarPlot function for the mean concentration of $\text{PM}_{2.5}$

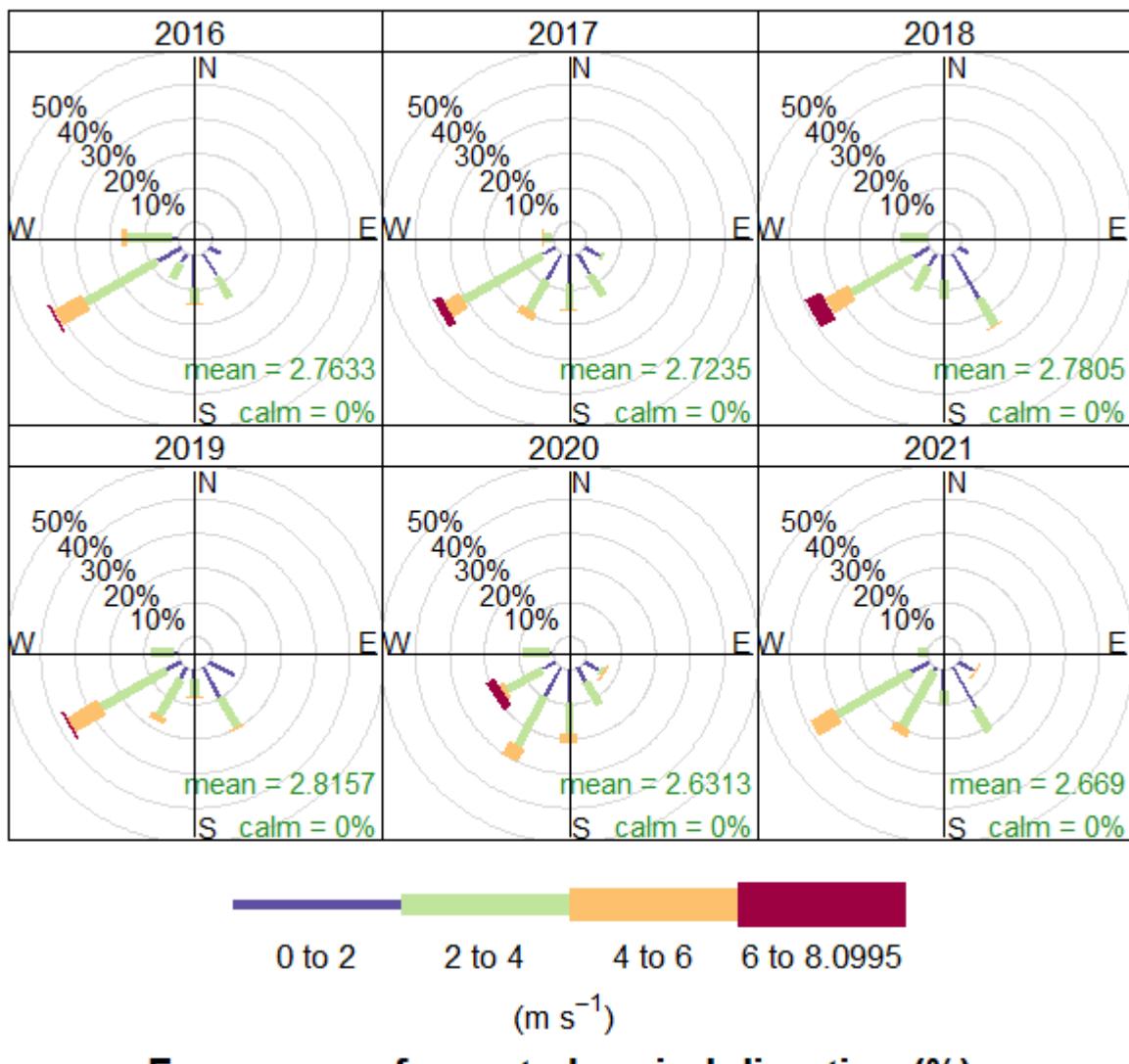


Figure 6.7. Analysis on wind speed/direction frequencies by year (2016-2021). Wind speeds are split into intervals shown by the scale in each panel. The grey circles show the 10-50 % frequencies.

6.3 Monthly variation in PM_{2.5} during lockdown

As previously discussed, the pollution levels experienced a significant acceleration in the early 2000s, with all districts surpassing an annual mean PM_{2.5} concentration of 35 $\mu\text{g m}^{-3}$ by the end of 2006. Subsequently, the high pollution levels gradually extended to the central MMR and its surrounding areas starting in early 2015 (Figure 6.8). Over time, all regions surrounding central MMR exceeded an annual mean PM_{2.5} concentration of 42 $\mu\text{g m}^{-3}$.

However, an interesting anomaly was observed in the year 2019, where pollution levels were lower compared to the preceding four years. This reduction in pollution can be attributed to prolonged monsoon periods and favorable weather conditions, which were also supported by the System of Air Quality and Weather Forecasting and Research (SAFAR) data (Times of India, 2020). It is worth noting that pollution levels are anticipated to increase in the post-lockdown years. However, surprisingly, we observed a significant and unexpected increase of up to 10% in five districts located during 2021 in the northeastern part of the MMR when compared with the BAU prediction. This suggests unanticipated changes in the local emission sources, such as industrial activities, vehicular traffic, or construction projects. These changes could have resulted in higher pollutant emissions than what was initially estimated by the BAU model.

This trend continued, with an insignificant decline of 10-25% observed between May and July. This aligns well with the study of Shehzad et al., (2021a) where the PM_{2.5} values showed an insignificant reduction up to 25% in June in Mumbai.

However, a strong significant reduction started to appear as low as 40% in August 2020, which corresponded well with the extended lockdown which only happened in Maharashtra. Between October and December 2020, insignificant differences were noticed between the predicted and actual observations, which could be attributed to the unlock phases. During January and February 2021, a substantial increase of 10-16% on average across all districts was observed in comparison with the BAU predicted values, coinciding with the lifting of all lockdown restrictions and a surge in COVID infections.

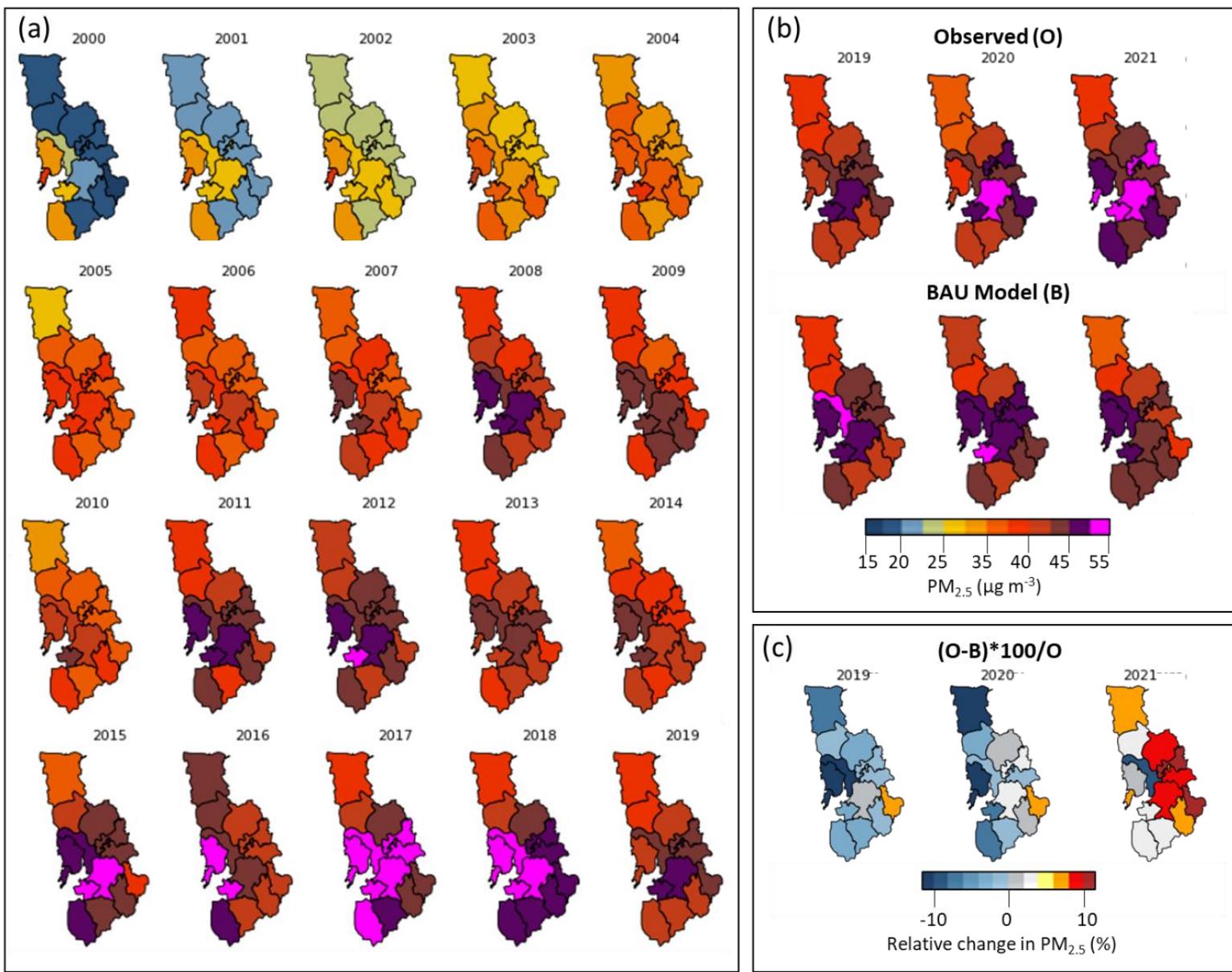


Figure 6.8: Trend in PM_{2.5} pollution across districts of MMR. Panel b shows the observed and BAU model output. Panel c shows the increase in observed PM_{2.5} in post-COVID year (2021).

In March 2020, prior to the implementation of the lockdown measures, there was a minor and insignificant decline of approximately 2% between the median observed value and the BAU prediction, indicating the accuracy of the model's forecast. However, during the month of April 2020, a significant decrease of about 47% was observed in comparison with the predicted values of business as usual, indicating the impact of the lockdown.¹⁰⁶ studies using satellite measurements support this, where they noticed a 41% decline during this lockdown month in Mumbai. The same month in the previous year shows less than 1% difference between BAU prediction and actual data, as 2019 was a BAU year.

From March to June 2021, no significant changes were observed in the observed and predicted values. Once again, a reduction of as low as 20-40% was observed between July and September 2021, which could be due to Phase 2 lockdown. We see that the decrease in pollution during 2020 and 2021 coincides well with lockdown durations (Figure 6.9).

We see no significant change in the rate of PM_{2.5} pollution during COVID year 2020 (Figure 6.10), despite the advent of stringent lockdowns. For all other years, except the 2019, anomalous year there is no significant change in rate. However, post COVID in 2021, the rate of pollution increased from -5.18 (BAU) to 6.24 $\mu\text{g m}^{-3}\text{a}^{-1}$. This suggests that pollution rate has increased more than expected (forecasted value) after the COVID lockdown. With respect to the weather conditions in 2021, the rate of PM_{2.5} decrease was predicted between -7.1% to -9.4%, however, -3% to 10.31% increase was observed in 2021. This could be due to an increase in economic activities that remained pending due to lockdown, highlighting the contribution of factories, road dust, stubble burning towards the dilapidated air quality, or background process changes in the atmosphere. This can also suggest that lockdown could be ineffective and outweigh the advantages in curbing pollution given the huge economy. These results are similar to the studies by National Air Quality Index (NAQI)¹⁰⁷ and PM_{2.5}¹⁰⁸ progressively increasing in post lockdown periods. Studies by P. Kumari & Toshniwal, (2020) also suggest that the air quality improvement in Indian cities is highly temporary as air quality continued to deteriorate once lockdown is lifted. However, in our study we reveal that the increase was more than the actual expected increase.

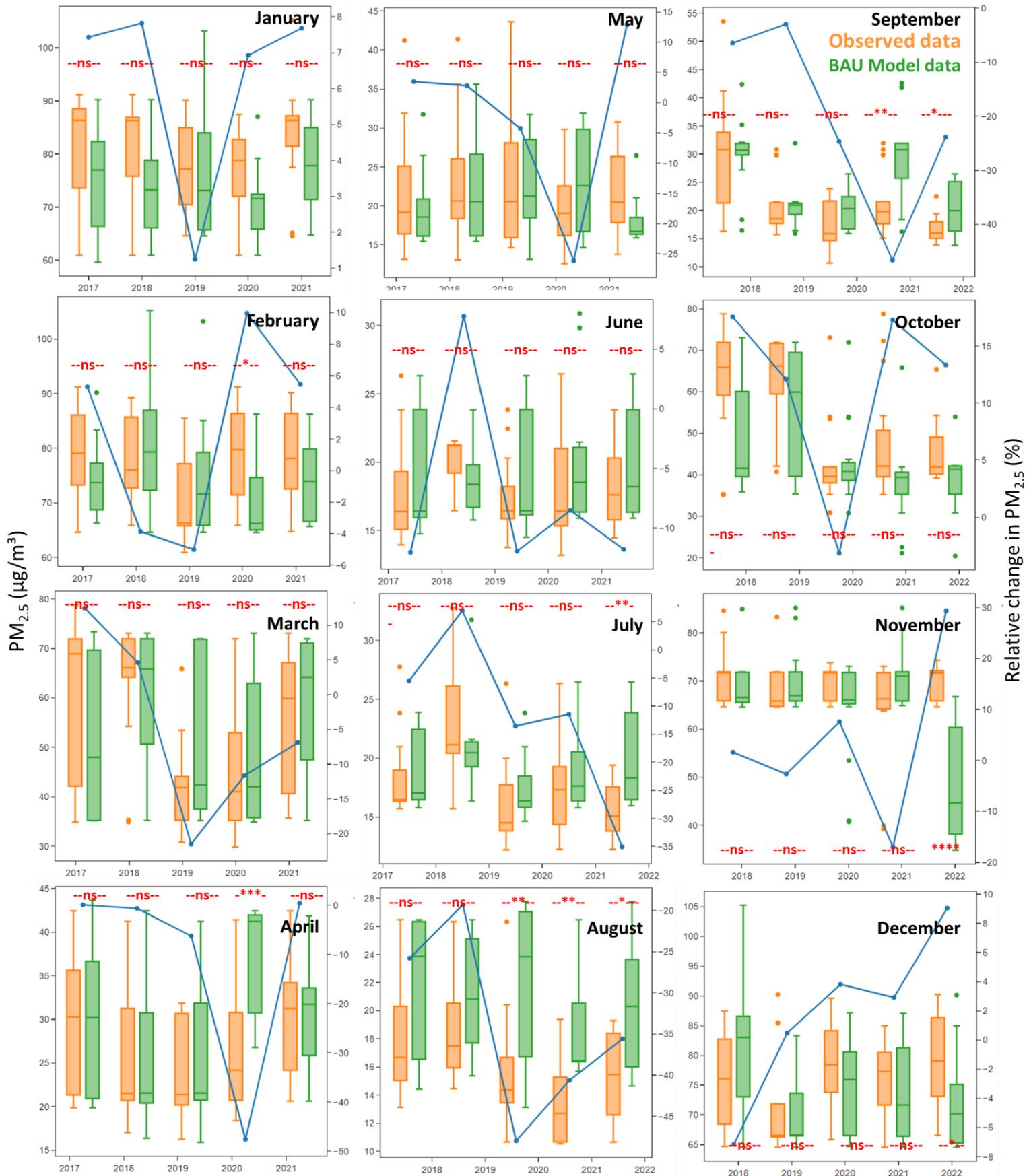


Figure 6.9: Comparison between actual PM_{2.5} and BAU Model value for 2017-2021.

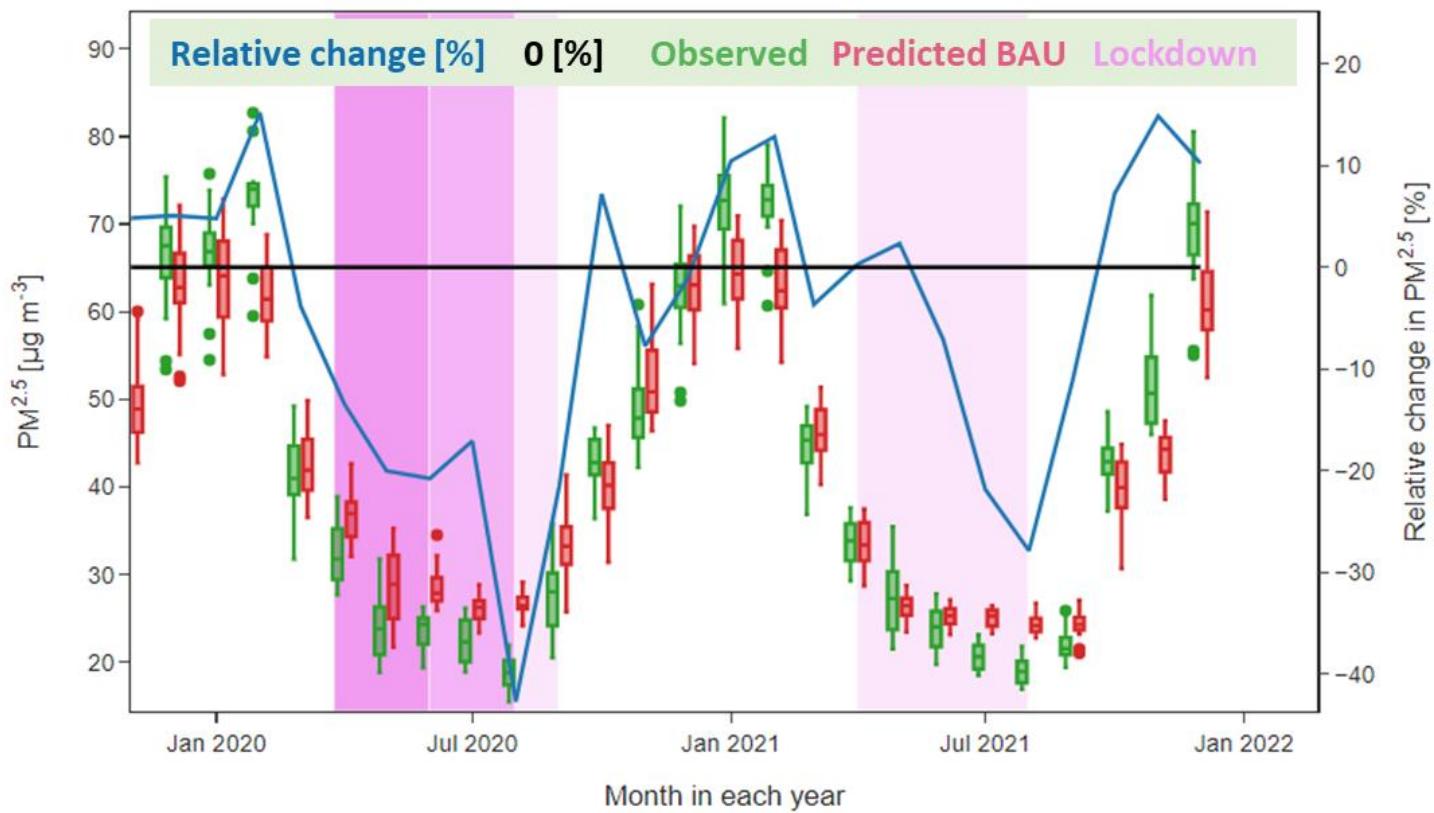


Figure 6.10: Trend in PM_{2.5} pollution across districts of MMR between 2020-2019. Boxes compare the average monthly mean with observed in green and predicted BAU in red. The black solid line shows 0% relative change between the observed and model and relative change for each month is given in blue solid line.

Chapter 7

Conclusions

We focused our study on the impact of COVID-19 and post-COVID-19 conditions on PM_{2.5} levels in the Mumbai Metropolitan Region (MMR) and aimed to predict business-as-usual (BAU) conditions during the COVID years of 2020 and 2021. We employed a random forest (RF) model trained on historical PM_{2.5} data and weather variables to predict PM_{2.5} concentrations. Our RF model demonstrated high accuracy and reliability in capturing the complex dynamics of PM_{2.5} levels during BAU periods, with a strong correlation (R-squared of 0.88) between predicted and observed PM_{2.5} levels. Our analysis of PM_{2.5} trends over the past two decades using the Seasonal-Trend decomposition using LOESS (STL) method revealed that PM_{2.5} levels increased significantly from the early 2000s until 2009, reflecting rapid urbanization and industrial growth. However, starting from 2009, the rate of increase in pollution became less pronounced, indicating a stabilization of PM_{2.5} levels. We also identified the influence of wind directions on PM_{2.5} concentrations, with winds from the east, south, and southeast contributing significantly to the increase in pollution, while winds from the southwest played a role in reducing pollution levels. During the COVID-19 pandemic, reduced industrial activity and decreased vehicular emissions resulted in lower pollutant levels, associated with stationary emission sources. However, as the region transitioned into the post-COVID phase, we observed a gradual increase in PM_{2.5} levels. This upward trend could be attributed to the resumption of economic activities, increased traffic, and industrial operations. While the pandemic provided a temporary respite from high pollution levels, the post-COVID period demonstrated the vulnerability of the MMR to pollution rebounds. The insights gained from our analysis can inform policymakers and stakeholders in developing effective strategies to mitigate pollution and ensure a healthier environment for the residents of the MMR and our study suggests that lockdown may not be a solution to tackle the increasing pollution.

In addition, we developed the first publicly available automated tool to obtain quality-assured air quality datasets by identifying and addressing these data issues. Notable variations in air quality metrics and invalidation of 15.5% of total data after data-cleaning suggest urgent need for data ratification. Current consecutive finding algorithms in CPCB's SMS alert system (CPCB, 2018) could be replaced with more robust metrics such as CoV. Future studies could consider incorporating mean, variance, autocorrelation, and diurnal pattern to evaluate the validity of large anomalous fluctuations. While CPCB cautions against using these records as absolute measures, it is imperative to allocate resources for proper functioning of CAAQMS in addition to instrument procurement. Urgent measures should be taken by data providers to rectify non-compliance with parameter reporting protocol.

Chapter 8

Annexure

Table 8.1. Parameters measured by CAAQMS stations in MMR region.

Place	Station	Parameter	Latitude	Longitude
Kalyan	Khadakpada, Kalyan - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.25292	73.142019
Mumbai	Bandra Kurla Complex, Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Toluene, Eth-Benzene, Xylene, MP-Xylene, AT, RH, WS, WD, BP, RF	19.053536	72.84643
Mumbai	Bandra, Mumbai - MPCB	---	19.0627	72.84614
Mumbai	Borivali East, Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.23241	72.86895
Mumbai	Borivali East, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.224333	72.865811
Mumbai	Chakala-Andheri East, Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Toluene, Eth-Benzene, Xylene, MP-Xylene, Temp, R H, WS, WD, BP, RF	19.11074	72.86084
Mumbai	Chhatrapati Shivaji Intl. Airport (T2), Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.10078	72.87462
Mumbai	Colaba, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	18.91	72.82
Mumbai	Deonar, Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Toluene, Eth-Benzene, Xylene, MP-Xylene, AT, RH, WS, WD, BP, RF	19.04946	72.923
Mumbai	Kandivali East, Mumbai - MPCB	NO, NO ₂ , NO _x , NH ₃	19.2058	72.8682
Mumbai	Kurla, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.0863	72.8888

Place	Station	Parameter	Latitude	Longitude
Mumbai	Malad West, Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, RH, WS, WD, BP	19.19709	72.82204
Mumbai	Mazgaon, Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Toluene, Eth-Benzene, Xylene, MP-Xylene, WS, WD	18.96702	72.84214
Mumbai	Mulund West, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.175	72.9419
Mumbai	Navy Nagar-Colaba, Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Toluene, Eth-Benzene, Xylene, MP-Xylene, AT, RH, WS, WD, BP, RF	18.897756	72.81332
Mumbai	Powai, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.1375	72.915056
Mumbai	Siddharth Nagar-Worli, Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , CO, O ₃ , AT, RH, WS, WD, BP, RF	19.000083	72.813993
Mumbai	Sion, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.047	72.8746
Mumbai	Vasai West, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, AT, RH, WS, WD, SR, BP	19.3832	72.8204
Mumbai	Vile Parle West, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.10861	72.83622
Mumbai	Worli, Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	18.993616	72.812811
Navi Mumbai	Airoli, Navi Mumbai - MPCB	---	19.1494	72.9986
Navi Mumbai	Nerul, Navi Mumbai - MPCB	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Eth-Benzene, MP-Xylene, AT, RH, WS, WD, SR, BP	19.008751	73.01662
Navi Mumbai	Sector-19A Nerul, Navi Mumbai - IITM	PM _{2.5} , PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Benzene, Toluene, Eth-Benzene, Xylene, MP-Xylene, AT, RH, WS, WD, BP, RF	19.044	73.0325
Thane	Pimpleshwar Mandir, Thane - MPCB	PM ₁₀ , NO, NO ₂ , NO _x , NH ₃ , SO ₂ , CO, O ₃ , Temp, WS, WD, SR	19.192056	72.958519

References

- (1) Kumar, R.; Joseph, A. E. Air Pollution Concentrations of PM2.5, PM10 and NO₂ at Ambient and Kerbsite and Their Correlation in Metro City - Mumbai. *Environ Monit Assess* **2006**, *119* (1–3), 191–199. [https://doi.org/10.1007/S10661-005-9022-7/METRICS](https://doi.org/10.1007/S10661-005-9022-7).
- (2) The World Bank. Health Impacts of Outdoor Air Pollution. **2003**.
- (3) Pandey, A.; Brauer, M.; Cropper, M. L.; Balakrishnan, K.; Mathur, P.; Dey, S.; Turkogulu, B.; Kumar, G. A.; Khare, M.; Beig, G.; Gupta, T.; Krishnankutty, R. P.; Causey, K.; Cohen, A. J.; Bhargava, S.; Aggarwal, A. N.; Agrawal, A.; Awasthi, S.; Bennett, F.; Bhagwat, S.; Bhanumati, P.; Burkart, K.; Chakma, J. K.; Chiles, T. C.; Chowdhury, S.; Christopher, D. J.; Dey, S.; Fisher, S.; Fraumeni, B.; Fuller, R.; Ghoshal, A. G.; Golechha, M. J.; Gupta, P. C.; Gupta, R.; Gupta, R.; Guttikunda, S.; Hanrahan, D.; Harikrishnan, S.; Jeemon, P.; Joshi, T. K.; Kant, R.; Kant, S.; Kaur, T.; Koul, P. A.; Kumar, P.; Kumar, R.; Larson, S. L.; Lodha, R.; Madhipatla, K. K.; Mahesh, P. A.; Malhotra, R.; Managi, S.; Martin, K.; Mathai, M.; Mathew, J. L.; Mehrotra, R.; Mohan, B. V. M.; Mohan, V.; Mukhopadhyay, S.; Mutreja, P.; Naik, N.; Nair, S.; Pandian, J. D.; Pant, P.; Perianayagam, A.; Prabhakaran, D.; Prabhakaran, P.; Rath, G. K.; Ravi, S.; Roy, A.; Sabde, Y. D.; Salvi, S.; Sambandam, S.; Sharma, B.; Sharma, M.; Sharma, S.; Sharma, R. S.; Shrivastava, A.; Singh, S.; Singh, V.; Smith, R.; Stanaway, J. D.; Taghian, G.; Tandon, N.; Thakur, J. S.; Thomas, N. J.; Toteja, G. S.; Varghese, C. M.; Venkataraman, C.; Venugopal, K. N.; Walker, K. D.; Watson, A. Y.; Wozniak, S.; Xavier, D.; Yadama, G. N.; Yadav, G.; Shukla, D. K.; Bekedam, H. J.; Reddy, K. S.; Guleria, R.; Vos, T.; Lim, S. S.; Dandona, R.; Kumar, S.; Kumar, P.; Landrigan, P. J.; Dandona, L. Health and Economic Impact of Air Pollution in the States of India: The

Global Burden of Disease Study 2019. *Lancet Planet Health* **2021**, 5 (1), e25–e38.

[https://doi.org/10.1016/S2542-5196\(20\)30298-9](https://doi.org/10.1016/S2542-5196(20)30298-9).

- (4) Schraufnagel, D. E.; Balmes, J. R.; Cowl, C. T.; De Matteis, S.; Jung, S. H.; Mortimer, K.; Perez-Padilla, R.; Rice, M. B.; Riojas-Rodriguez, H.; Sood, A.; Thurston, G. D.; To, T.; Vanker, A.; Wuebbles, D. J. Air Pollution and Noncommunicable Diseases: A Review by the Forum of International Respiratory Societies' Environmental Committee, Part 2: Air Pollution and Organ Systems. *Chest* **2019**, 155 (2), 417–426. <https://doi.org/10.1016/J.CHEST.2018.10.041>.
- (5) WHO. *Air quality database 2018*. <https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database/2018> (accessed 2023-06-10).
- (6) Chattopadhyay, A.; Shaw, S. Association Between Air Pollution and COVID-19 Pandemic: An Investigation in Mumbai, India. *Geohealth* **2021**, 5 (7). <https://doi.org/10.1029/2021GH000383>.
- (7) Desai, P. A. *Air pollution in city this winter significantly worse than last season / Mumbai news - Hindustan Times*. <https://www.hindustantimes.com/cities/mumbai-news/air-pollution-in-city-this-winter-significantly-worse-than-last-season-101675019641915.html> (accessed 2023-06-10).
- (8) Beig, G. *Why Mumbai's air became worse than Delhi's – and what can be done / The Indian Express*. <https://indianexpress.com/article/opinion/columns/why-mumbai-air-became-worse-than-delhis-what-can-be-done-8434585/> (accessed 2023-06-10).
- (9) WHO. *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed 2023-04-24).

- (10) The Indian Express. *Covid-19 India timeline: Looking back at pandemic-induced lockdown and how the country is coping with the crisis* / India News,The Indian Express. <https://indianexpress.com/article/india/covid-19-india-timeline-looking-back-at-pandemic-induced-lockdown-7241583/> (accessed 2023-04-24).
- (11) Nigam, R.; Pandya, K.; Luis, A. J.; Sengupta, R.; Kotha, M. Positive Effects of COVID-19 Lockdown on Air Quality of Industrial Cities (Ankleshwar and Vapi) of Western India. *Scientific Reports 2021 11:1* **2021**, 11 (1), 1–12. <https://doi.org/10.1038/s41598-021-83393-9>.
- (12) Nigam, R.; Pandya, K.; Luis, A. J.; Sengupta, R.; Kotha, M. Positive Effects of COVID-19 Lockdown on Air Quality of Industrial Cities (Ankleshwar and Vapi) of Western India. *Scientific Reports 2021 11:1* **2021**, 11 (1), 1–12. <https://doi.org/10.1038/s41598-021-83393-9>.
- (13) Times now news. *Unlock 3/ India unlock 3 guidelines: Interstate travel rules in the country.* <https://www.timesnownews.com/auto/features/article/india-unlock-3-guidelines-interstate-travel-rules-in-the-country/630430> (accessed 2023-05-08).
- (14) Kumar, S. Effect of Meteorological Parameters on Spread of COVID-19 in India and Air Quality during Lockdown. *Science of The Total Environment* **2020**, 745, 141021. <https://doi.org/10.1016/J.SCITOTENV.2020.141021>.
- (15) Kumari, S.; Lakhani, A.; Kumari, K. M. COVID-19 and Air Pollution in Indian Cities: World's Most Polluted Cities. *Aerosol Air Qual Res* **2020**, 20 (12), 2592–2603. <https://doi.org/10.4209/AAQR.2020.05.0262>.
- (16) Sharma, S.; Zhang, M.; Anshika; Gao, J.; Zhang, H.; Kota, S. H. Effect of Restricted Emissions during COVID-19 on Air Quality in India. *Science of The Total Environment* **2020**, 728, 138878. <https://doi.org/10.1016/J.SCITOTENV.2020.138878>.

- (17) Sarfraz, M.; Shehzad, K.; Shah, S. G. M. The Impact of COVID-19 as a Necessary Evil on Air Pollution in India during the Lockdown. *Environmental Pollution* **2020**, *266*, 115080. <https://doi.org/10.1016/J.ENVPOL.2020.115080>.
- (18) Malhotra, I.; Tayal, A. *IEEE Xplore Full-Text PDF*; 2021. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9441317> (accessed 2023-04-25).
- (19) Silva, A. C. T.; Branco, P. T. B. S.; Sousa, S. I. V. Impact of COVID-19 Pandemic on Air Quality: A Systematic Review. *Int J Environ Res Public Health* **2022**, *19* (4), 1950. <https://doi.org/10.3390/IJERPH19041950/S1>.
- (20) Zhang, K.; Liu, Z.; Zhang, X.; Li, Q.; Jensen, A.; Tan, W.; Huang, L.; Wang, Y.; De Gouw, J.; Li, L. Insights into the Significant Increase in Ozone during COVID-19 in a Typical Urban City of China. *Atmos Chem Phys* **2022**, *22* (7), 4853–4866. <https://doi.org/10.5194/ACP-22-4853-2022>.
- (21) Aburas, M. M.; Ho, Y. M.; Ramli, M. F.; Ash'aari, Z. H. The Simulation and Prediction of Spatio-Temporal Urban Growth Trends Using Cellular Automata Models: A Review. *International Journal of Applied Earth Observation and Geoinformation* **2016**, *52*, 380–389. <https://doi.org/10.1016/J.JAG.2016.07.007>.
- (22) News 18. *Mumbai Has Highest Vehicle Density in India Despite Lesser Cars Than Delhi.* <https://www.news18.com/news/auto/mumbai-has-highest-vehicle-density-in-india-despite-lesser-cars-than-delhi-2528843.html> (accessed 2023-04-25).
- (23) Van Donkelaar, A.; Hammer, M. S.; Bindle, L.; Brauer, M.; Brook, J. R.; Garay, M. J.; Hsu, N. C.; Kalashnikova, O. V.; Kahn, R. A.; Lee, C.; Levy, R. C.; Lyapustin, A.; Sayer, A. M.; Martin, R. V. Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty. *Environ Sci Technol* **2021**, *55* (22), 15287–15300.

https://doi.org/10.1021/ACS.EST.1C05309/ASSET/IMAGES/LARGE/ES1C05309_0007.jpeg.

- (24) Cooper, M. J.; Martin, R. V.; Hammer, M. S.; Levelt, P. F.; Veefkind, P.; Lamsal, L. N.; Krotkov, N. A.; Brook, J. R.; McLinden, C. A. Global Fine-Scale Changes in Ambient NO₂ during COVID-19 Lockdowns. *Nature* 2022 **601**:7893 **2022**, *601* (7893), 380–387. <https://doi.org/10.1038/s41586-021-04229-0>.
- (25) Bell, B.; Hersbach, H.; Simmons, A.; Berrisford, P.; Dahlgren, P.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Radu, R.; Schepers, D.; Soci, C.; Villaume, S.; Bidlot, J. R.; Haimberger, L.; Woollen, J.; Buontempo, C.; Thépaut, J. N. ERA5. *Quarterly Journal of the Royal Meteorological Society* **2021**, *147* (741), 4186–4227. <https://doi.org/10.1002/QJ.4174>.
- (26) ECMWF. <https://confluence.ecmwf.int/pages/viewpage.action?pageId=133262398> (accessed 2023-05-22).
- (27) Pandey, A.; Brauer, M.; Cropper, M. L.; Balakrishnan, K.; Mathur, P.; Dey, S.; Turkogulu, B.; Kumar, G. A.; Khare, M.; Beig, G.; Gupta, T.; Krishnankutty, R. P.; Causey, K.; Cohen, A. J.; Bhargava, S.; Aggarwal, A. N.; Agrawal, A.; Awasthi, S.; Bennett, F.; Bhagwat, S.; Bhanumati, P.; Burkart, K.; Chakma, J. K.; Chiles, T. C.; Chowdhury, S.; Christopher, D. J.; Dey, S.; Fisher, S.; Fraumeni, B.; Fuller, R.; Ghoshal, A. G.; Golechha, M. J.; Gupta, P. C.; Gupta, R.; Gupta, R.; Gupta, S.; Guttikunda, S.; Hanrahan, D.; Harikrishnan, S.; Jeemon, P.; Joshi, T. K.; Kant, R.; Kant, S.; Kaur, T.; Koul, P. A.; Kumar, P.; Kumar, R.; Larson, S. L.; Lodha, R.; Madhipatla, K. K.; Mahesh, P. A.; Malhotra, R.; Managi, S.; Martin, K.; Mathai, M.; Mathew, J. L.; Mehrotra, R.; Mohan, B. V. M.; Mohan, V.; Mukhopadhyay, S.; Mutreja, P.; Naik, N.; Nair, S.; Pandian, J. D.; Pant, P.; Perianayagam, A.; Prabhakaran, D.; Prabhakaran, P.; Rath, G.

K.; Ravi, S.; Roy, A.; Sabde, Y. D.; Salvi, S.; Sambandam, S.; Sharma, B.; Sharma, M.;
Sharma, S.; Sharma, R. S.; Shrivastava, A.; Singh, S.; Singh, V.; Smith, R.; Stanaway,
J. D.; Taghian, G.; Tandon, N.; Thakur, J. S.; Thomas, N. J.; Toteja, G. S.; Varghese, C.
M.; Venkataraman, C.; Venugopal, K. N.; Walker, K. D.; Watson, A. Y.; Wozniak, S.;
Xavier, D.; Yadama, G. N.; Yadav, G.; Shukla, D. K.; Bekedam, H. J.; Reddy, K. S.;
Guleria, R.; Vos, T.; Lim, S. S.; Dandona, R.; Kumar, S.; Kumar, P.; Landrigan, P. J.;
Dandona, L. Health and Economic Impact of Air Pollution in the States of India: The
Global Burden of Disease Study 2019. *Lancet Planet Health* **2021**, *5* (1), e25–e38.
[https://doi.org/10.1016/S2542-5196\(20\)30298-9](https://doi.org/10.1016/S2542-5196(20)30298-9).

- (28) Vohra, K.; Vodonos, A.; Schwartz, J.; Marais, E. A.; Sulprizio, M. P.; Mickley, L. J.
Global Mortality from Outdoor Fine Particle Pollution Generated by Fossil Fuel
Combustion: Results from GEOS-Chem. *Environ Res* **2021**, *195*, 110754.
<https://doi.org/10.1016/j.envres.2021.110754>.
- (29) Lelieveld, J.; Klingmüller, K.; Pozzer, A.; Burnett, R. T.; Haines, A.; Ramanathan, V.
Effects of Fossil Fuel and Total Anthropogenic Emission Removal on Public Health and
Climate. *Proc Natl Acad Sci U S A* **2019**, *116* (15), 7192–7197.
https://doi.org/10.1073/PNAS.1819989116/SUPPL_FILE/PNAS.1819989116.SD02.X
LSX.
- (30) Dutta, M.; Chatterjee, A. A Deep Insight into State-Level Aerosol Pollution in India:
Long-Term (2005–2019) Characteristics, Source Apportionment, and Future Projection
(2023). *Atmos Environ* **2022**, *289*. <https://doi.org/10.1016/j.atmosenv.2022.119312>.
- (31) Vohra, K.; Marais, E. A.; Bloss, W. J.; Schwartz, J.; Mickley, L. J.; Van Damme, M.;
Clarisso, L.; Coheur, P. F. Rapid Rise in Premature Mortality Due to Anthropogenic Air
Pollution in Fast-Growing Tropical Cities from 2005 to 2018. *Sci Adv* **2022**, *8* (14),

4435.

https://doi.org/10.1126/SCIADV.ABM4435/SUPPL_FILE/SCIADV.ABM4435_SM.PDF.

- (32) Ganguly, T.; Selvaraj, K. L.; Guttikunda, S. K. National Clean Air Programme (NCAP) for Indian Cities: Review and Outlook of Clean Air Action Plans. *Atmos Environ X* **2020**, 8, 100096. <https://doi.org/10.1016/J.AEAOA.2020.100096>.
- (33) CPCB. *Central Pollution Control Board (CPCB): National Ambient Air Quality Monitoring Series (NAAQMS)*. <https://cpcb.nic.in/publication-details.php?pid=OA==> (accessed 2023-03-09).
- (34) CPCB. *Annual Report 2020-21*; Delhi, 2021. <https://cpcb.nic.in/openpdffile.php?id=UmVwb3J0RmlsZXMuMTQwM18xNjU1MzU0NzKxX21lZGlhcGhvdG8xNjQ3MS5wZGY=> (accessed 2023-03-09).
- (35) Hancock, S.; Fiore, A. M.; Westervelt, D. M.; -, al; Liu, W.; Yang, Z.; Liu -, Q.; Stowell, J. D.; Bi, J.; Al-Hamdan, M. Z.; Suthar, G.; Singhal, R. P.; Khandelwal, S.; Kaul, N.; Parmar, V.; Singh, A. P. Four-Year Spatiotemporal Distribution & Analysis of PM2.5 and Its Precursor Air Pollutant SO₂, NO₂ & NH₃ and Their Impact on LST in Bengaluru City, India. *IOP Conf Ser Earth Environ Sci* **2022**, *1084* (1), 012036. <https://doi.org/10.1088/1755-1315/1084/1/012036>.
- (36) Mor, S.; Singh, T.; Bishnoi, N. R.; Bhukal, S.; Ravindra, K. Understanding Seasonal Variation in Ambient Air Quality and Its Relationship with Crop Residue Burning Activities in an Agrarian State of India. *Environmental Science and Pollution Research* **2022**, *29* (3), 4145–4158. <https://doi.org/10.1007/S11356-021-15631-6/FIGURES/4>.

- (37) Singh, B. P.; Kumar, P. Spatio-Temporal Variation in Fine Particulate Matter and Effect on Air Quality during the COVID-19 in New Delhi, India. *Urban Clim* **2021**, *40*, 101013. <https://doi.org/10.1016/J.UCLIM.2021.101013>.
- (38) Fatima, S.; Ahlawat, A.; Mishra, S. K.; Maheshwari, M.; Soni, V. K. Variations and Source Apportionment of PM_{2.5} and PM₁₀ Before and During COVID-19 Lockdown Phases in Delhi, India. *Mapan - Journal of Metrology Society of India* **2022**, *37* (4), 937–955. <https://doi.org/10.1007/S12647-021-00506-5/FIGURES/13>.
- (39) Prakash, J.; Choudhary, S.; Raliya, R.; Chadha, T. S.; Fang, J.; Biswas, P. Real-Time Source Apportionment of Fine Particle Inorganic and Organic Constituents at an Urban Site in Delhi City: An IoT-Based Approach. *Atmos Pollut Res* **2021**, *12* (11), 101206. <https://doi.org/10.1016/J.APR.2021.101206>.
- (40) Goyal, P.; Gulia, S.; Goyal, S. K.; Kumar, R. Assessment of the Effectiveness of Policy Interventions for Air Quality Control Regions in Delhi City. *Environmental Science and Pollution Research* **2019**, *26* (30), 30967–30979. <https://doi.org/10.1007/S11356-019-06236-1/FIGURES/8>.
- (41) Gulia, S.; Goyal, P.; Prakash, M.; Goyal, S. K.; Kumar, R. Policy Interventions and Their Impact on Air Quality in Delhi City — an Analysis of 17 Years of Data. *Water Air Soil Pollut* **2021**, *232* (11), 1–13. <https://doi.org/10.1007/S11270-021-05402-X/FIGURES/6>.
- (42) Vohra, K.; Marais, E. A.; Suckra, S.; Kramer, L.; Bloss, W. J.; Sahu, R.; Gaur, A.; Tripathi, S. N.; Van Damme, M.; Clarisse, L.; Coheur, P. F. Long-Term Trends in Air Quality in Major Cities in the UK and India: A View from Space. *Atmos Chem Phys* **2021**, *21* (8), 6275–6296. <https://doi.org/10.5194/ACP-21-6275-2021>.

- (43) Dey, S.; Purohit, B.; Balyan, P.; Dixit, K.; Bali, K.; Kumar, A.; Imam, F.; Chowdhury, S.; Ganguly, D.; Gargava, P.; Shukla, V. K. 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 Sens (Basel)* **2020**, *12* (23), 3872. <https://doi.org/10.3390/RS12233872>.
- (44) Maheshwarkar, P.; Sunder Raman, R. Population Exposure across Central India to PM_{2.5} Derived Using Remotely Sensed Products in a Three-Stage Statistical Model. *Scientific Reports* **2021** *11*:1 **2021**, *11* (1), 1–13. <https://doi.org/10.1038/s41598-020-79229-7>.
- (45) Prabhu, V.; Singh, P.; Kulkarni, P.; Sreekanth, V. Characteristics and Health Risk Assessment of Fine Particulate Matter and Surface Ozone: Results from Bengaluru, India. *Environ Monit Assess* **2022**, *194* (3), 1–17. <https://doi.org/10.1007/S10661-022-09852-6/TABLES/4>.
- (46) Sharma, K.; Garg, A.; Joshi, V.; Kumar, A. Assessment of Health Risks for Criteria Air Pollutants Present in 11 Non-Attainment Cities of Uttar Pradesh, India. *Human and Ecological Risk Assessment: An International Journal* **2022**, *29* (1), 103–122. <https://doi.org/10.1080/10807039.2022.2157242>.
- (47) CPCB. *CPCB: National Air Quality Index*; 2015. https://app.cpcbcr.com/CCR_docs/FINAL-REPORT_AQI_.pdf (accessed 2023-03-11).
- (48) Jaiswal, A. *India Announces National Air Quality Index to Increase Public Awareness / NRDC*. <https://www.nrdc.org/experts/anjali-jaiswal/india-announces-national-air-quality-index-increase-public-awareness> (accessed 2023-03-11).
- (49) Yusupov, M.; Yan, D.; Cordeiro, R. M.; -, al; Schaub, S.; Huber, R.; Ryu, Y.-H.; Min, S.-K. What Matters in Public Perception and Awareness of Air Quality? Quantitative

Assessment Using Internet Search Volume Data. *Environmental Research Letters* **2020**, *15* (9), 0940b4. <https://doi.org/10.1088/1748-9326/AB9FB0>.

- (50) CPCB. *Central Pollution Control Board*. <https://cpcb.nic.in/quality-assurance-quality-control/> (accessed 2023-03-11).
- (51) Roychowdhury, A.; Somvanshi, A. *Breathing Space: How to Track and Report Air Pollution Under the National Clean Air Programme*; 2020. <https://www.cseindia.org/content/downloadreports/9923> (accessed 2023-03-09).
- (52) Singh, V.; Singh, S.; Biswal, A.; Kesarkar, A. P.; Mor, S.; Ravindra, K. Diurnal and Temporal Changes in Air Pollution during COVID-19 Strict Lockdown over Different Regions of India. *Environmental Pollution* **2020**, *266*, 115368. <https://doi.org/10.1016/J.ENVPOL.2020.115368>.
- (53) Oksanen, E.; Pandey, V.; Pandey, A. K.; Keski-Saari, S.; Kontunen-Soppela, S.; Sharma, C. Impacts of Increasing Ozone on Indian Plants. *Environmental Pollution* **2013**, *177*, 189–200. <https://doi.org/10.1016/J.ENVPOL.2013.02.010>.
- (54) Cusworth, D. H.; Mickley, L. J.; Sulprizio, M. P.; Liu, T.; Marlier, M. E.; Defries, R. S.; Guttikunda, S. K.; Gupta, P. Quantifying the Influence of Agricultural Fires in Northwest India on Urban Air Pollution in Delhi, India. *Environmental Research Letters* **2018**, *13* (4), 044018. <https://doi.org/10.1088/1748-9326/AAB303>.
- (55) Datameet. <https://groups.google.com/g/datameet/c/TYHG43NJUx4> (accessed 2023-03-14).
- (56) Kumar, K.; Pande, B. P. Air Pollution Prediction with Machine Learning: A Case Study of Indian Cities. *International Journal of Environmental Science and Technology* **2022**, 1–16. <https://doi.org/10.1007/S13762-022-04241-5/TABLES/7>.

- (57) Sharma, D.; Mauzerall, D. Analysis of Air Pollution Data in India between 2015 and 2019. *Aerosol Air Qual Res* **2022**, *22* (2), 210204. <https://doi.org/10.4209/AAQR.210204>.
- (58) Greenstone, M.; Nilekani, J.; Pande, R.; Ryan, N.; Sudarshan, A.; Sugathan, A. *Lower Pollution, Longer Lives Life Expectancy Gains If India Reduced Particulate Matter Pollution*; 2015; Vol. 8. <https://www.jstor.org/stable/pdf/24481424.pdf> (accessed 2023-03-09).
- (59) Pant, P.; Lal, R. M.; Guttikunda, S. K.; Russell, A. G.; Nagpure, A. S.; Ramaswami, A.; Peltier, R. E. Monitoring Particulate Matter in India: Recent Trends and Future Outlook. *Air Qual Atmos Health* **2019**, *12* (1), 45–58. <https://doi.org/10.1007/S11869-018-0629-6/FIGURES/5>.
- (60) Sharma, C.; Mandal, T. K.; Singh, S.; Gupta, G.; Kulshrestha, M. J.; Johri, P.; Ranjan, A.; Upadhyaya, A. K.; Das, R. M.; Soni, D.; Mishra, S. K.; Muthusamy, S. K.; Sharma, S. K.; Singh, P.; Aggarwal, S. G.; Radhakrishnan, S. R.; Kumar, M. *Metrology for Atmospheric Environment*; Springer, Singapore, 2020. https://doi.org/10.1007/978-981-15-8872-3_14.
- (61) *CCR: Central Control Room for Air Quality Management - All India.* <https://app.cpcbcr.com/CCR/#/CAAQM-Dashboard-all/CAAQM-Landing> (accessed 2023-03-14).
- (62) Josh Brobst. Files for [Https://Stackoverflow.Com/q/60910962](https://stackoverflow.com/q/60910962) · GitHub. 2020. <https://gist.github.com/jbrobst/56c13bbbf9d97d187fea01ca62ea5112> (accessed 2023-03-31).
- (63) Sajjad Anwar. [India/India_district.Geojson at Master](#) · Geohacker/India · GitHub. March 14, 2014.

https://github.com/geohacker/india/blob/master/district/india_district.geojson

(accessed 2023-03-31).

- (64) Silver, B.; Reddington, C. L.; Arnold, S. R.; Spracklen, D. V. Substantial Changes in Air Pollution across China during 2015–2017. *Environmental Research Letters* **2018**, *13* (11), 114012. <https://doi.org/10.1088/1748-9326/AAE718>.
- (65) Zhai, S.; Jacob, D. J.; Wang, X.; Shen, L.; Li, K.; Zhang, Y.; Gui, K.; Zhao, T.; Liao, H. Fine Particulate Matter (PM_{2.5}) Trends in China, 2013-2018. Separating Contributions from Anthropogenic Emissions and Meteorology. *Atmos Chem Phys* **2019**, *19* (16), 11031–11041. <https://doi.org/10.5194/ACP-19-11031-2019>.
- (66) Rohde, R. A.; Muller, R. A. Air Pollution in China: Mapping of Concentrations and Sources. *PLoS One* **2015**, *10* (8), e0135749. <https://doi.org/10.1371/JOURNAL.PONE.0135749>.
- (67) Iglewicz, B.; Hoaglin, D. C. *Volume 16: How to Detect and Handle Outliers*", *The ASQC Basic References in Quality Control: Statistical Techniques*; 1993; Vol. 36. https://books.google.com/books/about/How_to_Detect_and_Handle_Outliers.html?id=siInAQAAIAAJ (accessed 2023-03-14).
- (68) Hanley, T. *PM 2.5 Continuous Monitoring Network and Method Update with Focus on Data Quality*; 2016. https://19january2017snapshot.epa.gov/sites/production/files/2016-10/documents/pm2.5_continuous_monitoring.pdf (accessed 2023-03-14).
- (69) Hart, D. *Met One Technical Bulletin Measurement Resolution Considerations for Monitoring PM 10 and PM 2.5 Simultaneously in Low Concentrations*; 2010. https://metone.com/wp-content/uploads/2019/04/pm_coarse_resolution.pdf (accessed 2023-03-14).

- (70) Yao, X.; Shaikh, K.; Lam, P. H.; Evans, G. J. Underestimation of Sulfate Concentration in PM_{2.5} Using a Semi-Continuous Particle Instrument Based on Ion Chromatography. *Journal of Environmental Monitoring* **2009**, *11* (6), 1292–1297. <https://doi.org/10.1039/B819630C>.
- (71) CPCB. *CPCB IT Division Protocol for Data Transmission from CAAQM Stations Existing as on Date*; 2015. https://app.cpcbccr.com/CCR_docs/Protocol_CAAQM.pdf (accessed 2023-03-14).
- (72) CPCB. *National Air Quality Index*; 2014. https://app.cpcbccr.com/CCR_docs/FINAL-REPORT_AQI_.pdf (accessed 2023-03-28).
- (73) Mathieu, E.; Ritchie, H.; Rodés-Guirao, L.; Appel, C.; Giattino, C.; Hasell, J.; Macdonald, B.; Dattani, S.; Beltekian, D.; Ortiz-Ospina, E.; Roser, M. Coronavirus Pandemic (COVID-19). *Our World in Data* **2020**.
- (74) Bhawan, P.; Nagar, A. *Guidelines for the Measurement of Ambient Air Pollutants VOLUME-II CENTRAL POLLUTION CONTROL BOARD Guidelines for Real Time Sampling & Analyses*; 2011. <http://www.indiaenvironmentportal.org.in/files/NAAQSManualVolumeII.pdf> (accessed 2023-03-19).
- (75) CPCB. *Compliance Reporting Protocols for Online Continuous Emission and Effluent Monitoring Systems (OCEMS)*; Delhi, 2018. <http://cpcb.nic.in/Online-Monitoring-Industrial-Emission-Effluent/> (accessed 2023-04-07).
- (76) van Zoest, V. M.; Stein, A.; Hoek, G. Outlier Detection in Urban Air Quality Sensor Networks. *Water Air Soil Pollut* **2018**, *229* (4). <https://doi.org/10.1007/S11270-018-3756-7>.

- (77) Cybersecurity and Infrastructure Security Agency. *Radio Frequency Interference Best Practices Guidebook*, February 2020; 2020.
https://www.cisa.gov/sites/default/files/publications/safecom-ncswic_rf_interference_best_practices_guidebook_2.7.20_-_final_508c.pdf (accessed 2023-03-19).
- (78) Hart, D. *Met One Technical Bulletin BAM-1020 Detection Limit*; 2013.
www.metone.com (accessed 2023-03-19).
- (79) United States Environmental Protection Agency. *Technical Note- Clarifications and Guidance on Shelter Temperature Guidance for Ambient Air Pollutant Methods* ; 2018.
https://www.epa.gov/sites/default/files/2018-03/documents/clarifications_on_shelter_temperature_for_gaseous_pollutant_methods_03_2018_0.pdf (accessed 2023-03-19).
- (80) Kimbrough, S.; Chris Owen, R.; Snyder, M.; Richmond-Bryant, J. NO to NO₂ Conversion Rate Analysis and Implications for Dispersion Model Chemistry Methods Using Las Vegas, Nevada near-Road Field Measurements. *Atmos Environ* **2017**, *165*, 23–34. <https://doi.org/10.1016/J.ATMOSENV.2017.06.027>.
- (81) Orru, H.; Andersson, C.; Ebi, K. L.; Langner, J.; Åström, C.; Forsberg, B. Impact of Climate Change on Ozone-Related Mortality and Morbidity in Europe. *European Respiratory Journal* **2013**, *41* (2), 285–294.
<https://doi.org/10.1183/09031936.00210411>.
- (82) Brown, P. E.; Izawa, Y.; Balakrishnan, K.; Fu, S. H.; Chakma, J.; Menon, G.; Dikshit, R.; Dhaliwal, R. S.; Rodriguez, P. S.; Huang, G.; Begum, R.; Hu, H.; D’souza, G.; Guleria, R.; Jha, P. Mortality Associated with Ambient PM_{2.5} Exposure in India:

Results from the Million Death Study. *Environ Health Perspect* **2022**, *130* (9), 097004-1-097004–097014. <https://doi.org/10.1289/EHP9538>.

- (83) Sarkar, M.; Das, A.; Mukhopadhyay, S. Assessing the Immediate Impact of COVID-19 Lockdown on the Air Quality of Kolkata and Howrah, West Bengal, India. *Environ Dev Sustain* **2021**, *23* (6), 8613–8642. <https://doi.org/10.1007/S10668-020-00985-7/FIGURES/16>.
- (84) Kumari, P.; Toshniwal, D. Impact of Lockdown Measures during COVID-19 on Air Quality—A Case Study of India. <https://doi.org/10.1080/09603123.2020.1778646> **2020**, *32* (3), 503–510. <https://doi.org/10.1080/09603123.2020.1778646>.
- (85) Shukla, S.; Khan, R.; Saxena, A.; Sekar, S.; Ali, E. F.; Shaheen, S. M. Appraisal of COVID-19 Lockdown and Unlocking Effects on the Air Quality of North India. *Environ Res* **2022**, *204*, 112107. <https://doi.org/10.1016/J.ENVRES.2021.112107>.
- (86) Mandal, J.; Samanta, S.; Chanda, A.; Halder, S. Effects of COVID-19 Pandemic on the Air Quality of Three Megacities in India. *Atmos Res* **2021**, *259*, 105659. <https://doi.org/10.1016/J.ATMOSRES.2021.105659>.
- (87) Jain, S.; Sharma, T. Social and Travel Lockdown Impact Considering Coronavirus Disease (COVID-19) on Air Quality in Megacities of India: Present Benefits, Future Challenges and Way Forward. *Aerosol Air Qual Res* **2020**, *20* (6), 1222–1236. <https://doi.org/10.4209/AAQR.2020.04.0171>.
- (88) Siddiqui, A.; Chauhan, P.; Halder, S.; Devadas, V.; Kumar, P. Effect of COVID-19-Induced Lockdown on NO₂ Pollution Using TROPOMI and Ground-Based CPCB Observations in Delhi NCR, India. *Environ Monit Assess* **2022**, *194* (10), 1–29. <https://doi.org/10.1007/S10661-022-10362-8/FIGURES/15>.

- (89) Police, S.; Sahu, S. K.; Tiwari, M.; Pandit, G. G. Chemical Composition and Source Apportionment of PM_{2.5} and PM_{2.5–10} in Trombay (Mumbai, India), a Coastal Industrial Area. *Particuology* **2018**, *37*, 143–153.
<https://doi.org/10.1016/J.PARTIC.2017.09.006>.
- (90) Behrer, A. P.; Lobell, D. Higher Levels of No-till Agriculture Associated with Lower PM_{2.5} in the Corn Belt. *Environmental Research Letters* **2022**, *17* (9), 094012.
<https://doi.org/10.1088/1748-9326/AC816F>.
- (91) Swati Dhingra; Maitreesh Ghatak. How Has Covid-19 Affected India's Economy? . *Economics observatory* **2022**, *17* (1), 1–22. <https://doi.org/10.1086/209911>.
- (92) Petetin, H.; Bowdalo, D.; Soret, A.; Guevara, M.; Jorba, O.; Serradell, K.; Pérez García-Pando, C. Meteorology-Normalized Impact of the COVID-19 Lockdown upon NO₂ Pollution in Spain. *Atmos Chem Phys* **2020**, *20* (18), 11119–11141.
<https://doi.org/10.5194/ACP-20-11119-2020>.
- (93) Kanniah, K. D.; Kamarul Zaman, N. A. F.; Kaskaoutis, D. G.; Latif, M. T. COVID-19's Impact on the Atmospheric Environment in the Southeast Asia Region. *Science of The Total Environment* **2020**, *736*, 139658.
<https://doi.org/10.1016/J.SCITOTENV.2020.139658>.
- (94) Collivignarelli, M. C.; Abbà, A.; Caccamo, F. M.; Bertanza, G.; Pedrazzani, R.; Baldi, M.; Ricciardi, P.; Carnevale Miino, M. Can Particulate Matter Be Identified as the Primary Cause of the Rapid Spread of CoViD-19 in Some Areas of Northern Italy? *Environmental Science and Pollution Research* **2021**, *28* (25), 33120–33132.
<https://doi.org/10.1007/S11356-021-12735-X/FIGURES/7>.

- (95) Fenech, S.; Aquilina, N. J.; Vella, R. COVID-19-Related Changes in NO₂ and O₃ Concentrations and Associated Health Effects in Malta. *Frontiers in Sustainable Cities* **2021**, 3, 1. [https://doi.org/10.3389/FRSC.2021.631280/BIBTEX](https://doi.org/10.3389/FRSC.2021.631280).
- (96) Zhang, Z.; Liu, Y.; Liu, H.; Hao, A.; Zhang, Z. The Impact of Lockdown on Nitrogen Dioxide (NO₂) over Central Asian Countries during the COVID-19 Pandemic. *Environmental Science and Pollution Research* **2022**, 29 (13), 18923–18931. <https://doi.org/10.1007/S11356-021-17140-Y/TABLES/3>.
- (97) Otmani, A.; Benchrif, A.; Tahri, M.; Bounakhla, M.; Chakir, E. M.; El Bouch, M.; Krombi, M. Impact of Covid-19 Lockdown on PM10, SO₂ and NO₂ Concentrations in Salé City (Morocco). *Science of The Total Environment* **2020**, 735, 139541. <https://doi.org/10.1016/J.SCITOTENV.2020.139541>.
- (98) Mahato, S.; Pal, S.; Ghosh, K. G. Effect of Lockdown amid COVID-19 Pandemic on Air Quality of the Megacity Delhi, India. *Science of The Total Environment* **2020**, 730, 139086. <https://doi.org/10.1016/J.SCITOTENV.2020.139086>.
- (99) Wang, Y.; Wen, Y.; Wang, Y.; Zhang, S.; Zhang, K. M.; Zheng, H.; Xing, J.; Wu, Y.; Hao, J. Four-Month Changes in Air Quality during and after the COVID-19 Lockdown in Six Megacities in China. *Environ Sci Technol Lett* **2020**, 7 (11), 802–808. https://doi.org/10.1021/ACS.ESTLETT.0C00605/ASSET/IMAGES/LARGE/EZ0C00605_0002.JPG.
- (100) BMC. <https://www.hindustantimes.com/cities/mumbai-news/mcap-pm2-5-and-pm10-levels-have-fallen-since-2015-101648052486217.html> (accessed 2023-05-22).
- (101) Zhang, K.; Huang, L.; Li, Q.; Huo, J.; Duan, Y.; Wang, Y.; Yaluk, E.; Wang, Y.; Fu, Q.; Li, L. Explicit Modeling of Isoprene Chemical Processing in Polluted Air Masses in Suburban Areas of the Yangtze River Delta Region: Radical Cycling and Formation of

- Ozone and Formaldehyde. *Atmos Chem Phys* **2021**, *21* (8), 5905–5917.
<https://doi.org/10.5194/ACP-21-5905-2021>.
- (102) David Carslaw. Openair. **2023**. <https://doi.org/10.1016/j.envsoft.2011.09.008>.
- (103) Pekney, N.; Davidson, C.; Zhou, L.; Hopke, P. Application of PSCF and CPF to PMF-Modeled Sources of PM_{2.5} in Pittsburgh.
<http://dx.doi.org/10.1080/02786820500543324> **2007**, *40* (10), 952–961.
<https://doi.org/10.1080/02786820500543324>.
- (104) Watson, A. Y.; Bates, R. R.; Kennedy, D. Atmospheric Transport and Dispersion of Air Pollutants Associated with Vehicular Emissions. **1988**.
- (105) *Mumbai Air Quality: In 2019, Mumbai breathed cleanest air in four years / Mumbai News - Times of India.* <https://timesofindia.indiatimes.com/city/mumbai/in-2019-mumbai-breathed-cleanest-air-in-four-years/articleshow/73063117.cms> (accessed 2023-05-22).
- (106) Shehzad, K.; Xiaoxing, L.; Ahmad, M.; Majeed, A.; Tariq, F.; Wahab, S. Does Air Pollution Upsurge in Megacities after Covid-19 Lockdown? A Spatial Approach. *Environ Res* **2021**, *197*, 111052. <https://doi.org/10.1016/J.ENVRES.2021.111052>.
- (107) Shehzad, K.; Xiaoxing, L.; Ahmad, M.; Majeed, A.; Tariq, F.; Wahab, S. Does Air Pollution Upsurge in Megacities after Covid-19 Lockdown? A Spatial Approach. *Environ Res* **2021**, *197*, 111052. <https://doi.org/10.1016/J.ENVRES.2021.111052>.
- (108) Sarkar, S. K.; Khan, M. M. H. Impact of COVID-19 on PM_{2.5} Pollution in Fastest-Growing Megacity Dhaka, Bangladesh. *Disaster Med Public Health Prep* **2021**, *1*. <https://doi.org/10.1017/DMP.2021.131>.

(109) Kumari, P.; Toshniwal, D. Impact of Lockdown on Air Quality over Major Cities across the Globe during COVID-19 Pandemic. *Urban Clim* **2020**, *34*, 100719.
[https://doi.org/10.1016/J.UCLIM.2020.100719.](https://doi.org/10.1016/J.UCLIM.2020.100719)