Final report

Evaluation of modelling techniques for air quality management in Delhi

Prepared for
Central Pollution Control Board

The Energy and Resources Institute
...towards global sustainable development

Contact details

R Suresh
T E R I
Darbari Seth Block
IHC Complex, Lodhi Road
New Delhi – 110 003
India

Tel. 2468 2100 or 2468 2111
E-mail sureshr@teri.res.in
Fax 2468 2144 or 2468 2145
Web www.teriin.org
India +91 • Delhi (0)11

Team

Advisor: Dr. Prodipto Ghosh, Distinguished Fellow, TERI
Project Investigator: Mr. R Suresh, Fellow, TERI
Co-project Investigator: Dr. Sumit Sharma, Director and Senior Fellow, TERI

Team Members
Mr. Shivang Agarwal, Project Associate, TERI
Md. Hafizur Rehman, Research Associate, TERI
Ms. Seema Kundu, Associate Fellow, TERI

Technical assistance
VITO, Belgium

Contents

Team Members iii

Technical assistance iii

VITO, Belgium iii

1. Background 9
2. Objectives 10
3. Methodology 11

3.1 Spatial mapping and identification of air pollution hotspots 11

3.2 Air quality forecasting and public advisories at the selected hotspot 12

3.3 Models for spatial mapping and forecasting 14

3.4 Study region 15

3.5 Activities in the project 23

4. Spatial mapping model for Delhi 35

4.1 Spatial model results based on air quality data 2015 to 2017 35

4.1.1 Spatial correlation analysis 2015-2017 36

4.1.2 Construction of RIO spatial mapping model (based on 2015-2017 dataset). 43

4.2 Spatial model results based on air quality data 2018 (Jan-Dec) 54

Calculation of a_i parameters applied for RIO 61

4.4 Conclusion: spatial modelling 79

5. OVL model: Forecasting setup 80

5.1 Forecasting setup (2015 – 2017) 80

5.2 Initial forecast validation results 84

5.3 Forecasting model setup (2018) 87

5.3.1 Meteo processing and correlation analysis 87

5.3.2 Forecast model setup 2018 91

5.3.4 Forecast model validation results (2018) 91

5.4 Forecasting model pilot testing for 2019 97

5.5 Time series analysis of forecasts 101

5.6 Conclusion 104

6. Training Program on spatial mapping and forecasting models 105

7. Conclusion 106

8. References 107

Annexure I: Monthly spatial maps for $PM_{2.5}$ from January 2018 to March 2019 108

Annexure II: Data cleaning procedures 138

Annexure III: Seasonal Spatial Map for hotspot identification 139

Annexure IV: Forecast for 2019 PM_{10}, $PM_{2.5}$, O_3, NO_2 144

List of Figures

Figure 1 Annual average PM_{10} concentrations in 2017 across Indian cities 9
Figure 2 Broad methodology for spatial mapping model..... 11
Figure 3 Broad methodology for application of forecasting model 13
Figure 4 Study-domain – Delhi city..... 17
Figure 5 Trends of air quality in Delhi city 17
Figure 6 Ambient air quality monitoring stations in Delhi. 19
Figure 7 Gridded study domain of Delhi. 20
Figure 8 Land-use data for the study domain (2016)..... 21
Figure 9 District-wise population density in Delhi..... 22
Figure 10 Road network in NCR including Delhi (2016) 23
Figure 11 Yearly average concentrations of NO_2, PM_{10} and $PM_{2.5}$ for 10 manual air quality monitoring locations across Delhi..... 25
Figure 12 Yearly average concentrations (2015-2018) of PM_{10}, $PM_{2.5}$, NO_2, and O_3 for 37 continuous air quality monitoring locations across Delhi..... 27
Figure 13 Hourly variation of ozone concentrations in 2017 at Dwarka monitoring station in Delhi 28
Figure 14 Seasonal variation of SO_2, NO_2, PM_{10} and $PM_{2.5}$ concentrations in different manual monitoring stations in Delhi (averaged for 2015-2017)..... 30
Figure 15 Monthly average SO_2 to NO2 ratio at different locations in Delhi (Averaged for 2015-2017) 30
Figure 16 Monthly average NO_2 to PM_{10} ratio at different locations in Delhi (Averaged for 2015-2017) 31
Figure 17 Monthly average NO_2 to $PM_{2.5}$ ratio at different locations in Delhi (Averaged for 2015-2017) 31
Figure 18 Monthly average PM2.5 to PM10 ratio at different locations in Delhi (Averaged for 2015-2017) 32
Figure 19 Visualization of the GHS dataset for Delhi - 2015 gridded at 250m. 37
Figure 20 Land cover dataset used for the analysis. 38
Figure 21 Coefficient of correlation (and its variation with the buffer radii) with GHS_2015_MEAN population data and altitude with PM_{10} values at different stations. 39
Figure 22 Coefficient of correlation of $PM_{2.5}$ concentrations with different landuse categories at different stations (averaged for different buffer sizes) 40
Figure 23 Coefficient of correlation (and its variation with the buffer radii) with GHS_2015_MEAN population, altitude, and landuse categories with $PM_{2.5}$ values at different stations. 40
Figure 24 Coefficient of correlation of ozone concentrations with different landuse categories at different stations (averaged for different buffer sizes) 41
Figure 25 Coefficient of correlation (and its variation with the buffer radii) with GHS_2015_MEAN population, altitude, and landuse categories with ozone values at different stations. 41
Figure 26 Coefficient of correlation of NO_2 concentrations with different landuse categories at different stations (averaged for different buffer sizes) 42

Figure 27 Coefficient of correlation (and its variation with the buffer radii) with GHS_2015_MEAN population, roads, and landuse categories with NO2 values at different stations. 42
Figure 28 Optimisation model runs to derive ai values for $PM_{2.5}$ 46
Figure 29 Optimisation model runs to derive ai values for PM_{10} 47
Figure 30 Optimisation model runs to derive ai values for ozone 48
Figure 31 Comparison of modelled (using RIO) and actual PM2.5 and PM_{10} concentrations for 2017 49
Figure 32 Comparison of modelled (using RIO) and actual ozone concentrations for 2017 . 50
Figure 33 Coefficient of correlation between daily modelled and actual pollutant concentrations at different stations (2017) 51
Figure 34 Spatial plot of modelled annual averaged PM_{10} concentrations using RIO for the year 2017 52
Figure 35 Spatial plot of modelled annual averaged $PM_{2.5}$ concentrations using RIO for the year 2017 53
Figure 36 Optimization of the weights of the land cover classes to improve the trend of the annual average $PM_{2.5}$ concentrations for the stations in Delhi against their beta-values (LC1-LC4). 60
Figure 37 Beta parameter values for the LC1 and LC2 scenarios 62
Figure 38 Comparison of modelled and observed $PM_{2.5}$ concentrations at different stations in Delhi under LC4 scenario (2018) 63
Figure 39 Plots under LC4 scenario showing the coefficient of correlation, bias and RMSE values between daily modelled and actual pollutant concentration for different stations in 2018 64
Figure 40 Ratios of a) bias to std. deviation and b) RMSE to std. deviation between the modelled and predicted values of PM2.5 at different stations in 2018. 65
Figure 41 Annual average PM2.5 map for 2018 using RIO with land cover information as spatial driver (LC4) on daily average basis. 67
Figure 42 Population density map of Delhi (2015) 68
Figure 43 Correlation of $PM_{2.5}$ concentrations with population density in the buffer around the monitoring stations. 69
Figure 44 Spatial map based on population density as the parameter 70
Figure 45 Performance of model using population density as the spatial driver – a) comparison of modelled and actual $PM_{2.5}$ concentrations, b) Plots of ratios of Bias and RMSE to standard deviations 70
Figure 46 Comparison between results of RIO model and ordinary krigging method 71
Figure 47 Model optimization run for the LC2 scenario for PM10 (2018) 72
Figure 48 Model validation for PM10 under LC2 scenario - Spatial 73
Figure 49 Temproal validation of daily modelled concetraitons with actual values at different stations under LC2 scenario 75
Figure 50 Ratios of a) bias to std. deviation and b) RMSE to std. deviation between the modelled and predicted values of PM10 at different stations in 2018. 76
Figure 51 Spatial maps for PM_{10} for 2018 under LC2 scenario 77
Figure 52 Correlation analysis for selected meteorological parameters in the ANN models. the horizontal axes show the input model parameter, the vertical axis shows the natural logarithm of the daily averaged PM_{10} concentrations. 82
Figure 53 Correlation analysis, averaged over the stations for each pollutant 83

Figure 54 NMB value between observed concentration and forecasts for different pollutant at various stations (2017) 84
Figure 55 RMSE between observed concentration and forecasts for different stations and pollutants (2017) 85
Figure 56 Coefficient of correlation between observed concentration and forecasted prediction value 85
Figure 57 Coefficient of correlation between actual and predicted daily average concentrations in the forecast horizon of 5 days 86
Figure 58 Normalized mean bias for daily average concentrations - raw ANN as a function of forecast horizon, each dot represents an average over the stations. 87
Figure 59 Correlation analysis for selected meteo parameters, concentrations and fire events in the ANN models. 89
Figure 60 Correlation analysis averaged over all stations for day+1 averaged concentrations based on observations from 01/01/2018 – 31/12/2018. 90
Figure 61 Daily number of fire events around Delhi from January 2017 until March 2019 90
Figure 62 Comparison of relative RMSE averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2. 92
Figure 63 Relative RMSE per station for OPAQ Delhi model for year 2018 with RTC, no Resampling, forecast horizon day+1. Blue: NO2; Red: O3; Green: PM10; Purple: PM2.5 93
Figure 64 Comparison of normalized mean bias averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2. Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS. 93
Figure 65 Average normalized mean bias per station for OPAQ Delhi model for year 2018 with RTC, no Resampling, forecast horizon day+1. Blue: NO2; Red: O3; Green: PM10; Purple: PM2.5. 94
Figure 66 Comparison of R2 averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3/ and (d) NO2. Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS 94
Figure 67 Average R2 per station for OPAQ Delhi model for year 2018 with RTC, no Resampling, forecast horizon day+1. Blue: NO2; Red: O3; Green: PM10; Purple: PM2.5. 95
Figure 68 Comparison of FCF averaged over all stations as a function of forecast horizon for (a) PM2.5 and (b) PM10. Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS. 95
Figure 69 Average FCF of stations in Delhi for year 2018 for forecast horizon day+1. Blue: PM10; Red: PM2.5 96
Figure 70 Comparison of FFA averaged over all stations as a function of forecast horizon for (a) $PM_{2.5}$ and (b) PM_{10}. *Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS* 97
Figure 71 Average FFA per station for OPAQ Delhi model for year 2018 for forecast horizon day+1. Blue: PM10; Red: PM2.5 97
Figure 72 Comparison of relative RMSE considering all the models averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2. 99
Figure 73 Comparison of normalized mean bias considering all models averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2 100

Figure 74 Comparison of R2 considering all models averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2 101
Figure 75 Comparison of daily observed PM2.5 average concentrations (yellow) at Ashok Vihar station with , predicted concentrations day+1 no RTC (blue), predicted concentrations day +1 with RTC (green). 102
Figure 76 Comparison of daily observed PM2.5 average concentrations (yellow) at Major Dhyan Chand station with , predicted concentrations day+1 no RTC (blue), predicted concentrations day +1 with RTC (green). 102
Figure 77 Comparison of daily observed PM2.5 average concentrations (yellow) at Wazirpur station with , predicted concentrations day+1 no RTC (blue), predicted concentrations day +1 with RTC (green). 103
Figure 78 Comparison of daily observed PM2.5 average concentrations (yellow) at NSIT Dwarka station with , predicted concentrations day+1 no RTC (blue), predicted concentrations day +1 with RTC (green). 103

List of Table

Table 1 Correlation analysis for SO_2 33
Table 2 Correlation analysis for NO_2 33
Table 3 Correlation analysis for PM_{10} 34
Table 4 Correlation analysis for $PM_{2.5}$ 34
Table 5 Summary of stations used in RIO 2018 35
Table 6 Details of monitoring stations and their ambient air pollutant concentrations (annually averaged) 36
Table 7 RIO classes (based on land cover) 44
Table 8 Initial estimates of ai derived from available sectoral emission inventory estimates 44
Table 9 Optimized ai values for different landuse categories 45
Table 10 Annual Emission inventory of pollutants (kt/yr) in Delhi and NCR for 2016 54
Table 11 Broad landuse categories used in the model 55
Table 12 Area under each of the landuses in Delhi 55
Table 13 Area under clubbed landuse category in Delhi 56
Table 14 Initial ai values derived using sectoral emission estimates of Delhi city 56
Table 15 Description of scenarios used for model optimisation purposes 57
Table 16 Average concentration levels for different land uses in Delhi for 2018 78
Table 17 Parameters extracted over the Delhi domain from FNL data archive 81
Table 18 Parameters for model input vector 81
Table 19 Parameters extracted over the Delhi domain from FNL data archive 88
Table 20 Parameters for the model input vector 88
Table 21 Number of stations with sufficient data in 2018 and 2019 98

1. Background

Rapidly growing economy and migration of population to urban areas in India are the main drivers for degradation of ambient air quality (AAQ) and associated health impacts in the urban areas. According to WHO, Indian cities like Delhi, Gwalior, Raipur and Lucknow are among the top 100 most polluted cities in the world in terms of ambient PM10 concentrations (WHO, 2014). Various health studies conducted in India have evidently shown that the deteriorating ambient air quality results in significant health impacts (Cropper et al., 1997; Guttikunda and Goel, 2000; Kandlikar and Ramachandran, 2000; Dholakia et al., 2013). Figure 1 shows that air pollutant levels in most Indian cities where air quality monitoring is carried out, are above the prescribed national ambient air quality standards.

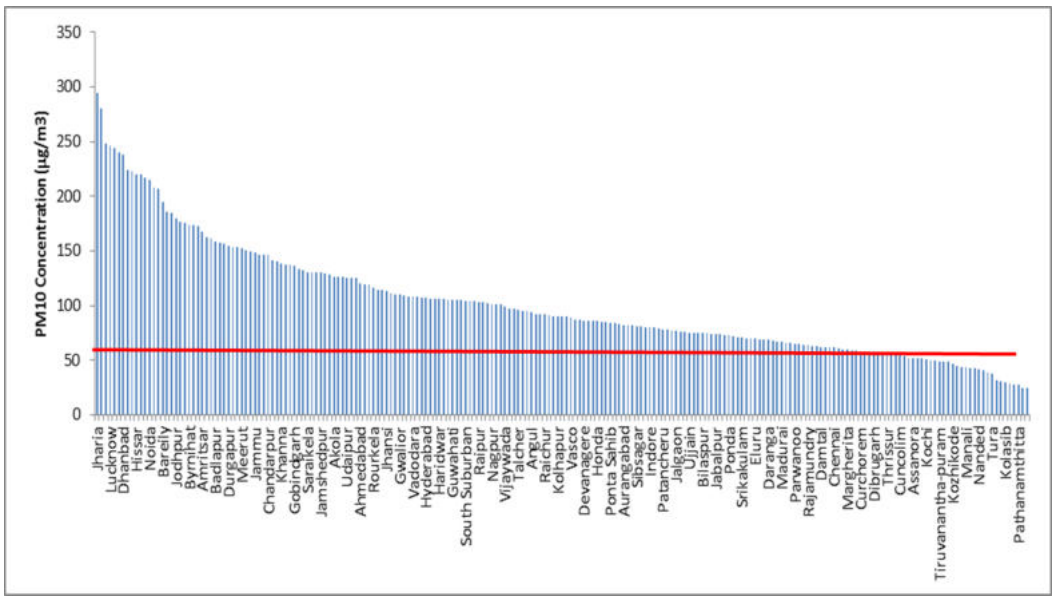


Figure 1 Annual average PM_{10} concentrations in 2017 across Indian cities
Source: NAMP, CPCB

In response to the growing concern around air pollution, the Government of India has taken several measures for control. Government has launched the National Clean Air Program in 2019, with an objective of reducing PM concentrations by 20-30% by 2024. Government has been aggressively promoting clean cooking in rural households through three major schemes: - Pradhan Mantri Ujjwala Yojana, Unnat Chulha Abhiyan and National Biogas and Manure Management Programme. In the transport sector some of the recent initiatives at the National level include introduction of BS IV norms in 2017 and rescheduling implementation of BS VI norms in the country from 2020. In addition, stricter stack emission norms have been developed for power plants, which are yet to be implemented.

Despite these efforts, the air pollution has remained high in several cities. The problem is more acute in specific zones of the cities due to presence of local sources. Presently, there is no mechanism in place to assess the spatial distributions of pollutant levels in cities and moreover, there is no forecasting system in most cities to know and prepare for an upcoming episode of high air pollution. Considering the enormity of the problem, it

becomes very important to understand the spatial distribution of air quality in the city and forecasts for coming days to identify effective measures for mitigation and reducing the exposure of residents to high levels of pollution levels in urban areas. In order to minimize the exposure to high concentrations of air pollution one should know the present air quality levels and should also be aware about probable air quality scenario of the next few days in advance. Evidently, this will require the knowledge of todays and forecasted air pollutant levels at all places in a city at a very high resolution.

Monitoring stations in different cities provide air pollutant concentrations only for specific locations in the city and rest of the city remains devoid of the air quality data. Moreover, there is no mechanism to forecast air quality in most of the cities to know in advance the air quality levels. TERI proposed this project which aimed to map air quality level in a typical city in India, like Delhi, and identify air quality hotspots using datasets generated by existing monitoring stations. The project also aimed to generate forecasts of air quality at the monitoring stations in Delhi for next few days , which can act as an early warning system for the citizens and relevant authorities to take adaptive measures for reducing the exposure to severely high episodic air pollutant concentrations.

Based on the data of measurements carried out at the stations in Delhi, spatial air quality maps can be developed with limited resources and can be made available on-line in almost real time basis. These maps not only provide air quality values at all locations in the city, but can also be useful in assessing the exposure of pollutants to residing population. This can lead to enhanced sensitization of general public and other stakeholders and will also help the policy makers to adjudge the impact of any interventions they make on spatial air quality levels in the whole city. Similarly, forecasting of air quality for next few days has now become possible through application of forecasting techniques. Forecasting models can predict the air quality in a region reasonably well so that mitigative/adaptive measures can be taken in advance to tackle the problem. Based on the proposal submitted to CPCB, TERI has been awarded the project of "Evaluation of modelling techniques for air quality management in Delhi". This project aimed at testing the air quality models for generating spatial maps of air quality and short term air quality forecasts. TERI (India) and VITO (Belgium) have signed a MoU to work jointly on a number of activities related to environmental protection. TERI with support by VITO have tested these models in Indian cities for generation of spatial air quality maps and air quality forecasts.

2. Objectives

The main objectives of the study are:

- Spatial mapping of pollutants in the city using existing monitoring datasets
- Identification of air pollution hotspots in the city
- Air quality forecasting and public advisories at the selected hotspots

3. Methodology

3.1 Spatial mapping and identification of air pollution hotspots

The purpose of spatial mapping is to estimate concentrations of air pollutants in areas devoid of any air quality monitors, based on concentrations measured at other locations. The overall approach followed in the project is presented in Figure 2.

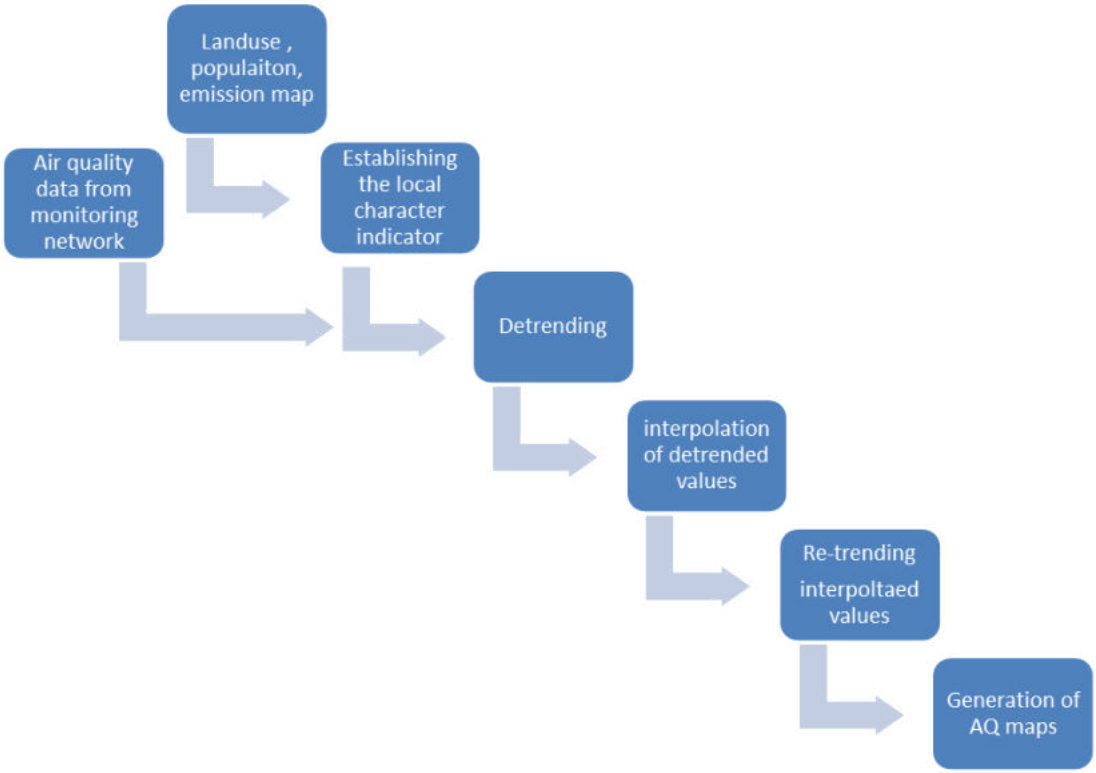


Figure 2 Broad methodology for spatial mapping model

The basic assumption of RIO model (developed by VITO) is that the concentrations of air pollutants are somewhat dependent on the land use pattern or any other parameter like population distribution, emission distribution, etc. in the city. The basic way to define categorize land use pattern (which is used as the parameter to affect air quality in this study) is – urban and rural. However, the land use pattern was categorized in much more detail by (Janssen et al. 2008). The land use pattern is divided into different categories (such as Continuous urban fabric, Discontinuous urban fabric , Industrial and commercial units , Road and rail networks , Airports , Construction sites , Agricultural areas , Forest and semi natural areas, Wetlands and water-bodies etc.). The model primarily interpolates values between the stations, however, interpolation of air quality requires spatial homogeneity. A β-indicator is assigned to each of the eleven landuse categories based on their possible influences on air quality. Hourly and daily air pollutant concentrations for pollutants like PM_{10}, $PM_{2.5}$, NO_x, and Ozone are collected from CPCB and DPCC network for the whole year 2015-17, and 2018 (Jan-Dec). These β-indicators were first optimised to derive

maximum correlation with the pollutant concentrations. RIO is an interpolation model that can be classified as a detrended Kriging model. In a first step the local characteristics of the air pollution sampling values is removed in detrending procedure. Subsequently the site independent data is interpolated by an ordinary Kriging scheme. Finally in retrending step a local bias is added to the Kriging interpolation results. Finalised β-values are then used to detrend concentrations of air pollutants to achieve spatial homogeneity for interpolation. The detrended (x_{det}) value can be obtained using equation 1.

$$X_{det} = (x-[x])(\sigma_{ref}/\sigma_{\beta}) + [x] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Where [x] is the mean concentration of the pollutant, σ_{ref} is the standard deviation of reference value which is chosen arbitrarily (as random sampling eliminates bias by giving all individuals an equal chance to be chosen, it removes systematic bias) and σ_{β} is the standard deviation of the β indicator. The reasons for choosing arbitrarily are

These detrended values are then spatially homogenous and can be interpolated using improved Kriging interpolation technique. Through this technique, RIO construct correlation function using time averaged values for interpolation model. Finally, in the interpolated grids, retrending is done in the same way to account for local character of the grids. The interpolated concentrations are validated for some of the locations by 'leaving one out' approach. Once the model is validated, spatial maps of air pollutant concentrations are generated. In this study, maps are prepared for different pollutants under different model (β-optimisation) scenarios for two different years 2015-17 and 2018 for the city of Delhi. Dataset for the year 2018 was separately assessed and tested, as there were many stations which were added to the network in this year. The results of RIO model can be overlaid on Google maps/earth for clean identification of areas with deteriorated air quality. This will help in identifying highly polluted areas/ hotspots in Delhi in different seasons of the year. $PM_{2.5}$ and PM_{10} being the pollutant of concern, the focus has been kept on spatial mapping of these two pollutants.

3.2 Air quality forecasting and public advisories at the selected hotspot

The main objective of this exercise is to predict the future pollutant concentration at the selected hotspots based on reliable input variables. However, input variables are usually multidimensional and the functional relationship with the target value to be predicted is usually unknown; and most likely non-linear, this makes it difficult to use the traditional parametric regression technique.

Artificial neural network (ANN) has to be designed for each monitoring site to fit a function between chosen inputs and target values. ANN systems are computing systems that are inspired by, but not necessarily identical to the biological neural network. Such system "learns' to perform tasks by considering examples or previous data sets. Large amounts of historical dataset would be required for this purpose. A part of the dataset would be used to train the neural network and the rest of the dataset would be used to validate the accuracy of the NN. The broad methodology for development of forecasting model for the city is shown in Figure 3.

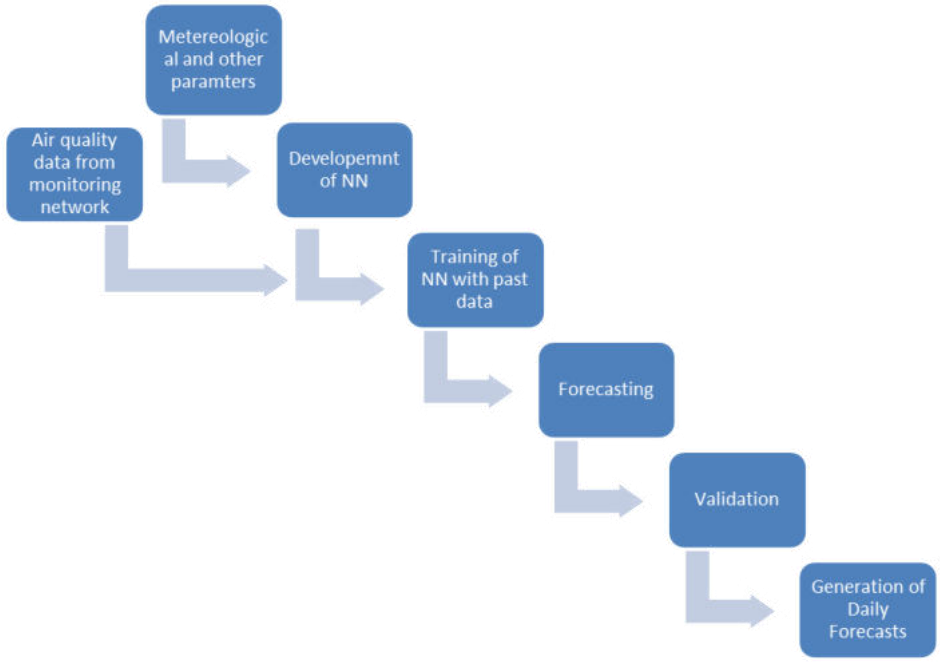


Figure 3 Broad methodology for application of forecasting model

The most crucial part of this exercise would be selection of appropriate input variables. For example, data on meteorological parameters like wind speed, wind direction, temperature etc. will be required for estimating concentrations of pollutants at a particular location. In addition, pollutants are also dispersed vertically in the atmosphere. Thus the concentration of pollutant on ground (target value in this case) is also dependent on the stability of the atmosphere.

All these parameters are tested in case of Delhi for prediction of pollutant concentrations. Past data of pollutants is correlated with these meteorological parameters and the parameters which showed the best performance are used for forecasting of air quality at specific stations in Delhi. Best performances are judged by hit and trial method. First few variables are chosen; air quality was forecasted and validated. This step was performed number of times to achieve the final set of variables. Daily forecasted values of the meteorological parameters are collected from global; simulation products (European Centre for Medium-range Weather Forecasts (ECMWF)). These forecasts for meteorological variables are used to compute the forecasted values of PM using neural network approach.

PM (day 1) = f (input parameters like PM (past days) , Forecasted met parameters (Boundary layer, wind speed , direction, temperature , day of week) day 1

The forecasted values are first validated against the observations. Relevant performance metrics like coefficient of correlation, Root Mean Square Error (RMSE), index of agreement are used to assess the performance of the forecasting model. Once fully validated, the model can start providing forecasts on a daily basis.

3.3 Models for spatial mapping and forecasting

Spatial Mapping

Air quality is measured at a few selected locations in Indian cities; attempts have been made across the world to map the concentrations of air pollutants in the space between these stations, based on the information provided by monitoring instruments at the stations. At a regional scale, variations in air pollutant concentrations are mainly due to meteorological conditions (Tombette and Sportisse, 2007; Mensink et al.,2007), while at urban scale it is deeply influenced by local emission sources (Vautard et al., 2007). Inverse distance weighting (IDW) is perhaps the most basic technique for spatial mapping (Isaaks and Srivastava, 1989). The IDW methodology allocates weight to the monitoring stations. This weight is determined by the user based on the distance of, and between the monitoring sites. The approximate concentration of pollutants is then obtained at the required site by the sum of products of pollutant concentrations at monitoring locations and weights associated with them. Although the technique is simple and reasonable, it is has weaknesses in the assumption. First of all, the decay function (weightage) is based on the power law of the distance between the monitoring location and the site of the air pollution interpolation. This power is chosen on an ad-hoc basis rather than character of the phenomenon. Secondly, monitoring locations close to one another are bound to have similar air quality monitoring values. A good interpolation model must somehow account for this factor. However, IDW only takes into account the distance between the monitoring locations and the interpolation site. Also, IDW assumes spatial homogeneity and does not incorporate the possibility of spatial trend (topography).

For improvement, a number of different spatial mapping models for air quality have been developed around the world. Ross et al. (2006) developed regression equation to predict fine particulate matter using parameters like urbanization, land use pattern, population density and industrialisation. In fact, most of the spatial maps of air quality have been developed in some regions of the world in past using interpolation models (Ross et al., 2007; Arain et al., 2007; Denby et al., 2005) based on different variants of the Kriging technique along with the use of landuse datasets. Broadly based on Kriging technique, the RIO model was developed and implemented in Belgium for spatial mapping of pollutants like PM, NOx and Ozone (Janssen, 2008). The RIO model 'detrends' or removes the local characteristics from the measured values, which results in spatial homogeneity. These values are then interpolated using the Kriging technique and finally 'retrended' on the local map. The RIO model incorporates a wide range of land use pattern, much beyond the basic rural and urban divisions. As a result, the model has shown to predict much more accurate results as compared to IDW and ordinary Kriging interpolation (Janssen, 2008).

Forecasting

Air quality forecasting is of importance to both policymakers and general public. A number of methodologies have been evolved in order to forecast air quality – Persistence, Climatology, Criteria, CART,

Regression, Neural Network, 3-D air quality models, and Phenomenological demonstrated the use of chemical transport modelling system for prediction of air quality. The technique is

data intensive and requires collection of variables of meteorology, activity data, emissions of pollutants from sources like industries, transport, biomass burning and so on. This methodology has an advantage that not just it predicts air quality in areas without air quality monitors but also helps in improving the understanding chemistry behind pollutant formation and dispersion processes. However, for accurate prediction of air quality, this technique requires high resolution high quality input data and high speed computing infrastructure for calculations, making it limited to be used.

Air quality can also be forecasted by developing regression equations in which pollutant concentration is calculated from several dependent variables like air quality in past days, temperature, wind speed and so on. Although development of regression model requires knowledge of the behaviour of the pollutants, variables affecting the pollutants, and execution techniques of regression models, the accuracy of regression model is highly dependent on the accuracy of the input variables.

(Artificial) neural networks (NN) have been used in many parts of the world for the purpose of forecasting. They are known for their high capabilities to make regressive approximations of nonlinear functions in spaces. Gardner and Dorling (1998) shows the different research studies using NN as tools for forecasting of air pollutants like ozone, sulphur dioxide and carbon monoxide. While deterministic tools demand huge sets of input data and computational facilities, NN are less data, time, and resource intensive, however, they are not useful in explaining the science behind the forecasted values and are site-specific. There are several studies that have proved NN to be an effective method for prediction of PM concentrations (Nagendra et al, 2005), Perez and Reyes, 2002, Lu et al. 2003, Kukkonen et al. 2003, Ordieres et al. 2005).

Based on the measurements carried out at the existing monitoring stations in a city, air quality forecasts can be made using the NN technique. Forecasts can be made available for next 2-3 days with limited resources and can be made available on-line. Forecasts of the pollutant can provide early warnings to the residents and regulators to take proactive adaptive actions. This can also lead to enhanced sensitization of general public and other stakeholders and will also help the policy makers to plan policies that can take care of pollution levels during the high episodes of pollution.

Hooyberghs et al, 2005 shows the successful implementation of the forecasting model based on neural networks for daily averaged PM_{10} concentrations in Belgium. TERI has built its capacity in running the OVL model developed by VITO and propose to implement and test this model in the Delhi city for generation of daily forecasts of air pollutant concentrations. The forecasting model can generate daily forecasts of air quality in Indian cities with limited resources, by making use of existing network of air quality measurement. These forecasts can also be generated by making use of chemical transport models which are based on emission inventories and are time and resource intensive. Alternatively, this model makes use of existing monitoring infrastructure to predict the air quality with satisfactory accuracy.

3.4 Study region

Delhi city has been selected as the study domain (Figure 4). Delhi being the capital city accommodates a huge population base of about 16.8 million. The city population has grown at a higher rate than the national average mainly because of extensive migration towards the

city. The city houses populations of different classes, varying from high income groups to people living in slums. The registered vehicular population in the city has grown from about 3 million in 1998 to about 10 million in 2017; with more than 60% of them are two-wheelers. Other than the vehicles registered in the Delhi city, there is a large in and out movement of vehicles from its surrounding towns like Gurgaon, Faridabad, Sonepat, Ghaziabad, and Noida. Delhi also has few power plants based on gaseous fuels. Other than these, there are power plants based on coal and gaseous fuels in the surrounding towns. Though, less in Delhi, there are frequent power cuts in other parts of NCR which leads to usage of standby power sources like diesel generators. There not many polluting industries in Delhi as most of them have been shifted to the outside regions. However, there is significant industrial fuel consumption in neighboring districts of Panipat, Merrut, Faridabad, Ghaziabad etc. With limited standards of NOx, NMVOCs and CO, emissions are released uncontrolled from many of these sources. Biomass burning is another big source which contributes to deterioration of air quality in Delhi. Not many in Delhi, but several households in NCR use biomass as fuel for cooking purposes. Moreover, post harvesting, agricultural residues are being burnt in the fields to prepare the lands for the next crop. Road dust, construction activities, and refuse burning are the other contributors to pollutant levels in the city.

Figure 4 Study-domain – Delhi city
Source: Google maps

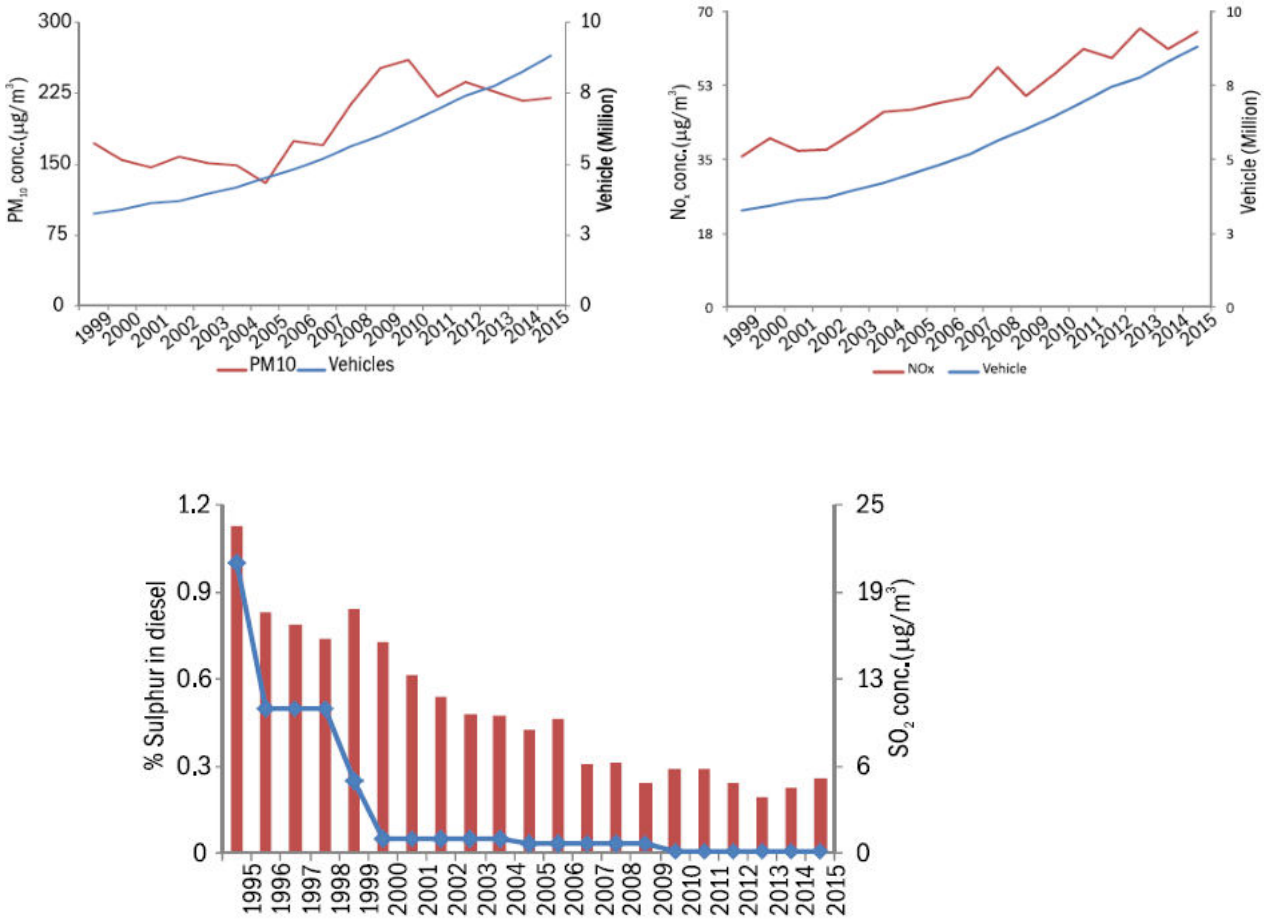


Figure 5 Trends of air quality in Delhi city

As seen from Figure 5, the number of registered vehicles in Delhi has grown tremendously over the past decade, crossing over nine million in 2015. However, a decline in PM_{10} concentration has been witnessed during 1999 to 2005, due to implementation of several control measures in transport and other sectors. However, after that a sharp increase in PM_{10} concentrations has been observed. The concentration of NOx has consistently gone up and shows a good correlation with increase in number of vehicles. On the other hand, the levels of SO_2 were found to be decreasing with decrease in sulphur content in automotive fuels.

The population growth, urbanization, growing activity levels in Delhi have resulted in corresponding increase in energy demand and henceforth air pollution from both stationary and mobile sources. As evident from the regular AAQ monitoring conducted in Delhi by CPCB and DPCC, the air pollution levels in Delhi have consistently been exceeding the AAQ standards especially in terms of $PM_{2.5}$ and PM_{10} concentrations. The major sources that are responsible for deteriorating ambient air quality in Delhi have been assessed through various source apportionment studies. The CPCB source apportionment study conducted in 2010 reported re-suspended dust (45%) as the largest contributor of PM_{10} in the city followed by waste burning (14%), transport (14%), DG sets (9%),industries (8%) and domestic cooking (7%). The study conducted by IIT Kanpur in 2015 reported that in winter season the three major sources of $PM_{2.5}$ were secondary particles (30%), biomass burning (26%) and transport (25%); while in summers the major source was found to be soil and road dust (28%) followed by coal and fly ash (26%) and secondary particles (15%). The latest study conducted by TERI & ARAI (2018) shows that in winters, industries (30%), transport (28%), and biomass (15%) are the major contributors to $PM_{2.5}$ concentrations.

AAQ Monitoring network in Delhi

AAQ monitoring in Delhi is conducted under National Air Monitoring Programme (NAMP) through various organizations which includes Central Pollution Control Board (CPCB), Delhi Pollution Control Committee (DPCC), NEERI, IMD, NDMC and Delhi Cantonment Board. Among all the monitoring stations manual air pollution monitoring is carried out at 10 stations in Delhi namely Sarojini Nagar, Chandni Chowk, Mayapuri Industrial Area, Pitampura, Shahadra, Shahzada Bagh, Nizamuddin, Janakpuri, Siri Fort, and at ITO as traffic intersection station. Out of 10 manual monitoring stations, three stations are managed by NEERI and the remaining seven stations managed by CPCB. While the continuous ambient air quality monitoring (CAAQM) stations works at 11 locations which includes AnandVihar, Civil Lines, DCE, Dilshad Garden, Dwarka, IGI Airport, ITO, MandirMarg, Punjabi Bagh, R.K. Puram, and Shadipur. The air quality monitoring stations managed by DPCC includes 6 locations i.e. Civil lines, Punjabi Bagh, Mandir Marg, Anand Vihar ISBT, IGI Airport, and R.K. Puram. As per the latest information provided by CPCB, under NAMP currently there are 38 continuous air quality monitoring stations in Delhi, out of which 24 stations are operated by DPCC, 7 stations each operated by IMD and CPCB.

Additionally, under System of Air Quality and Weather Forecasting and Research (SAFAR) of Indian Institute of Tropical Meteorology (IITM), Pune, AAQ monitoring is conducted at 8 monitoring stations in Delhi on real time basis. In addition to this, the U.S. Embassy and Consulates' monitor airborne fine particulate matter i.e $PM_{2.5}$ in the compounds of the Embassy and Consulates.

The ambient air quality monitoring stations operated by different agencies are shown in Figure 6.

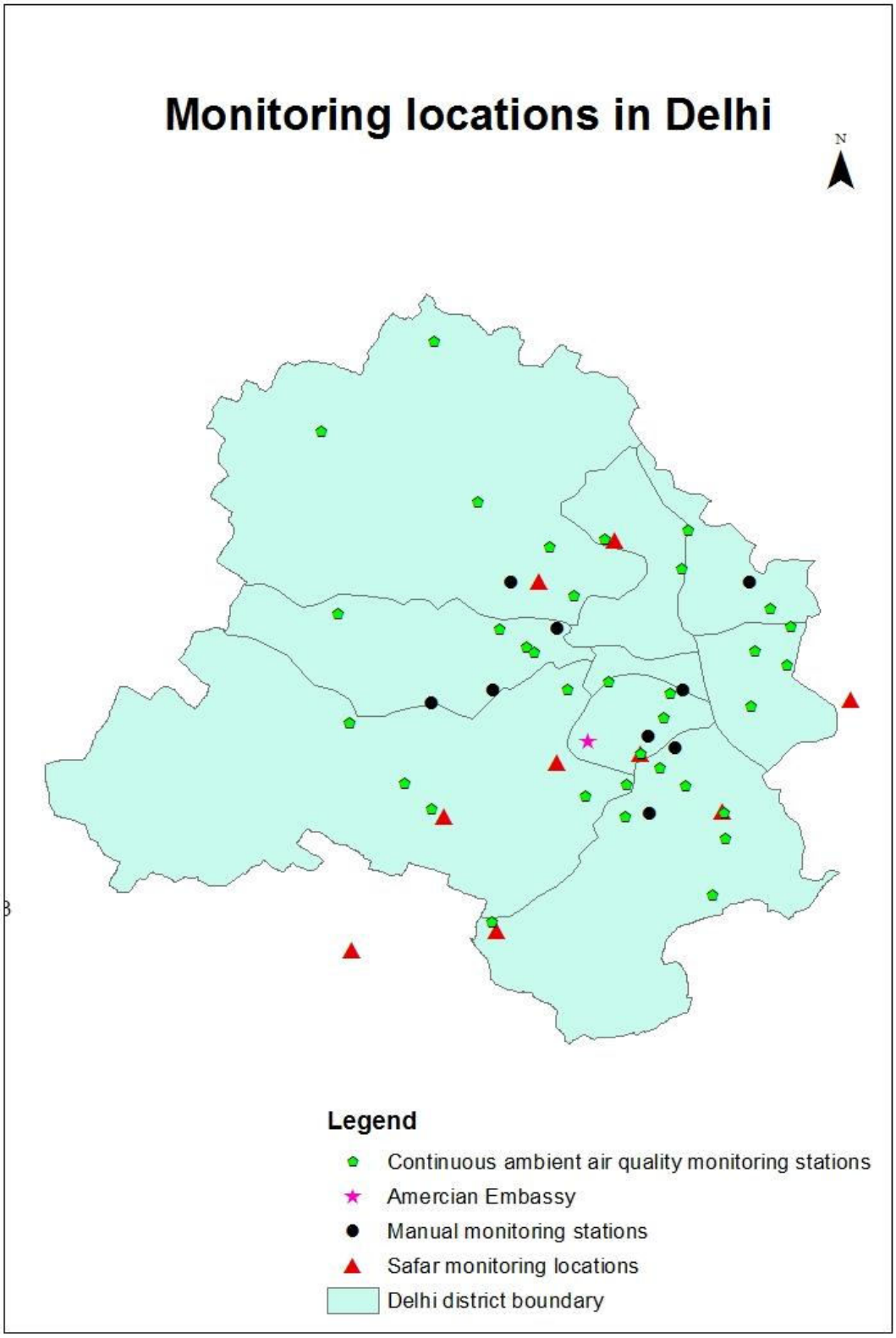


Figure 6 Ambient air quality monitoring stations in Delhi.

GIS mapping of study domain

ARC-GIS software has been used to divide the study domain into grids of 1x1 km^2. Figure 7 shows the gridded study domain, which is used to generate spatial air quality maps. Grids with existing monitoring stations of CPCB and DPCC have also been identified, which will be used for further analysis using RIO model.

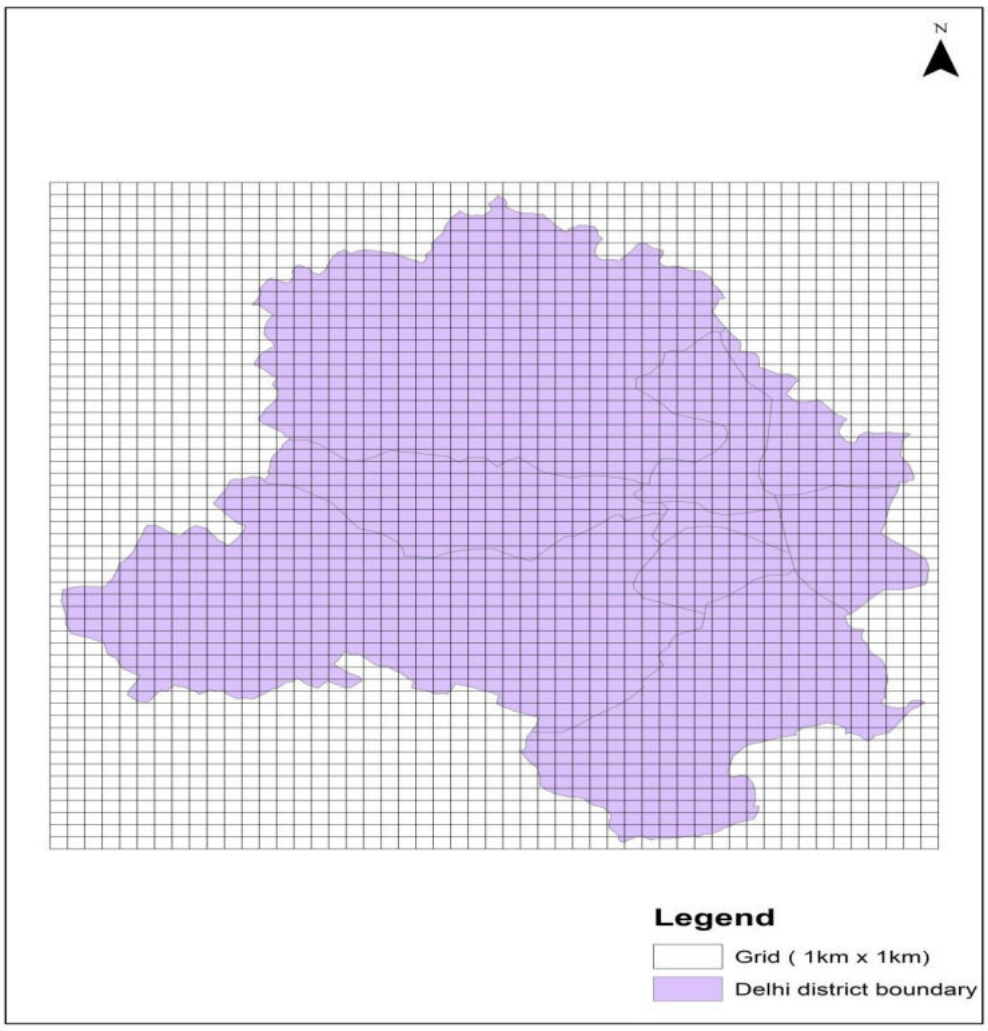


Figure 7 Gridded study domain of Delhi.

The land-use data for Delhi has been digitized for different land-use categories, using google earth. The category-wise digitized land-use of Delhi is shown in Figure 8. Evidently, central part of Delhi is more residential with some pockets of commercial interests. Outer parts of the north-western Delhi are still under the cultivable land category. There are specific industrial areas which are located in different parts of the city. The monitoring stations in the city are located in different land-use categories and form the basis of interpolation of values within different grids.

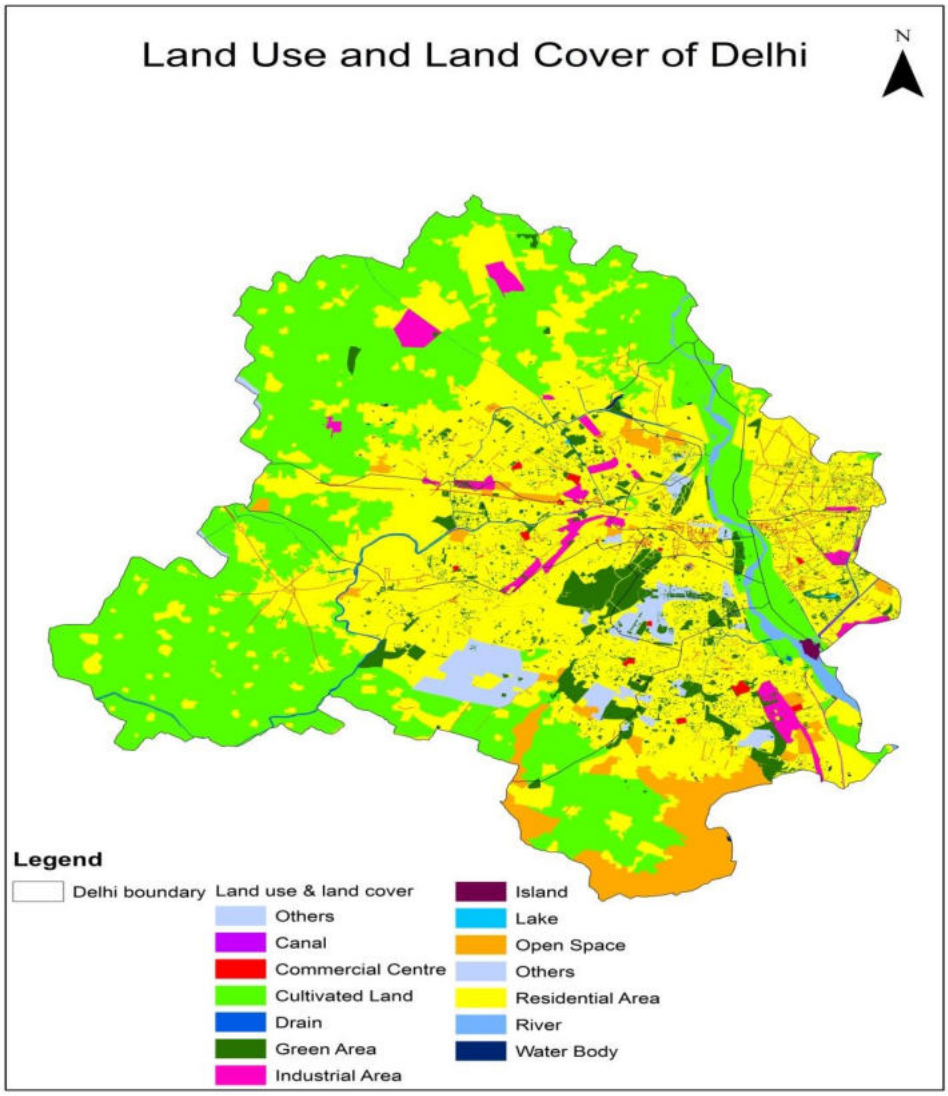


Figure 8 Land-use data for the study domain (2016)
Data source : Google earth

Population is unevenly distributed in the city of Delhi. The varying population density in the city is shown in Figure 9. Evidently, the central eastern parts of Delhi are most densely populated areas. However, outer regions in north-western Delhi show the lowest population densities. Population density is a driver for many activities, which lead to generation of emissions. Gridded population density will be estimated and the parameter is tested to possibly improve the prediction of the interpolated values by the RIO model.

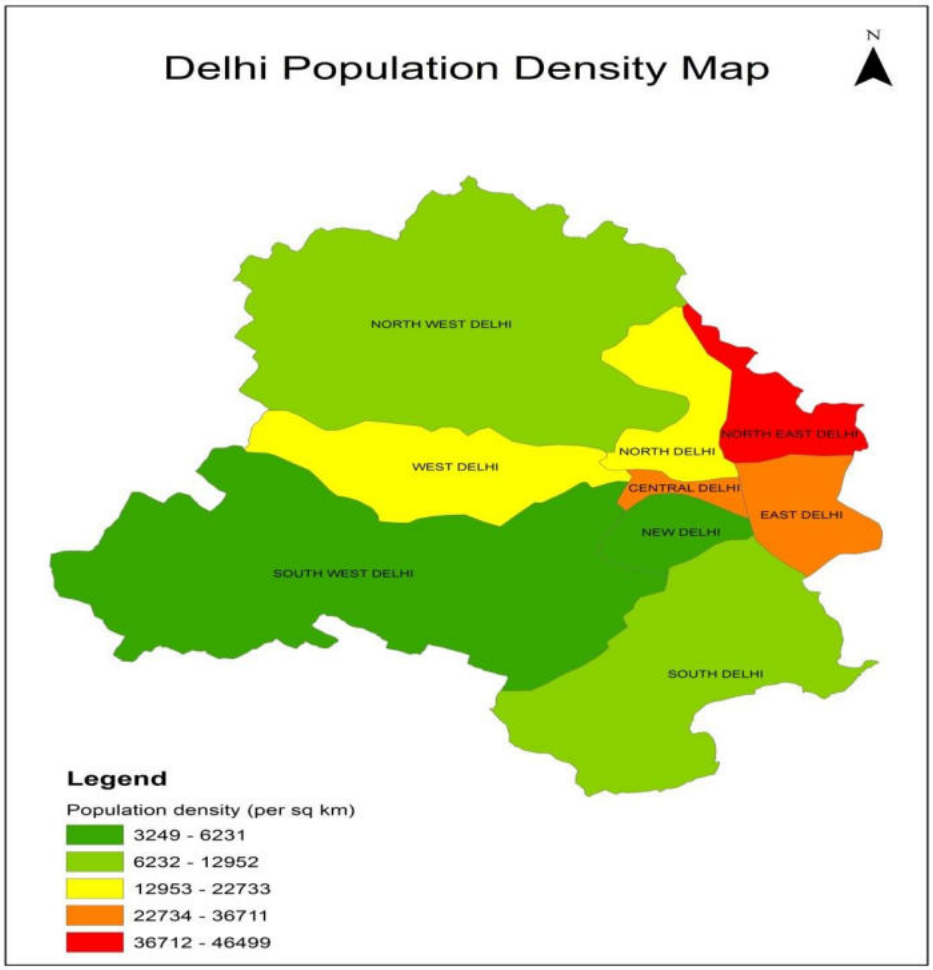


Figure 9 District-wise population density in Delhi

TERI already possesses a database of roads in the NCR and the same has been used for gridding and assessing the monitoring locations influenced by vehicles in the city. The roads have been identified in 3 categories – arterial, sub-arterial and minor roads (figure 10).

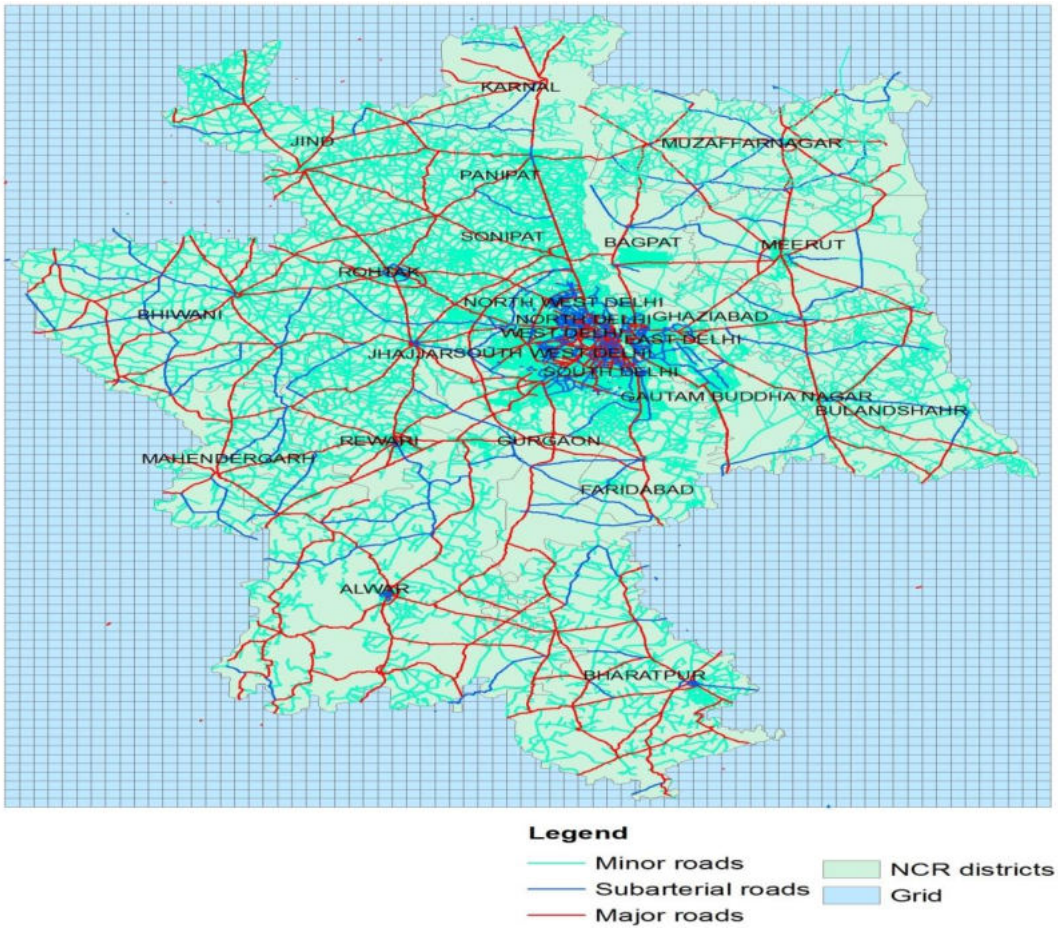


Figure 10 Road network in NCR including Delhi (2016)
Source: Digitized by TERI

3.5 Activities in the project

Project initiation

After the receipt of work order from CPCB, the project was initiated in the project management system in TERI. A project team of environmental engineers, scientists, and analysts was formed and an internal project initiation meeting was carried out to familiarize the team members with the project.

Data collection

Manual monitoring stations 24-hourly average ambient air quality data was collected from CPCB for the year 2015-2017 and later for the period Jan-Dec 2018, for the pollutants PM_{10}, $PM_{2.5}$ and NO_2. Data was also collected for the continuous monitoring stations (1-hourly average air quality data) for the same period for pollutants PM_{10}, $PM_{2.5}$, NO_2 and O_3, from CPCB's website (http://www.cpcb.gov.in/CAAQM/frmUserAvgReportCriteria.aspx) . It was observed that data is missing for several hours for various monitoring stations and a data gap report was prepared.

Data analysis

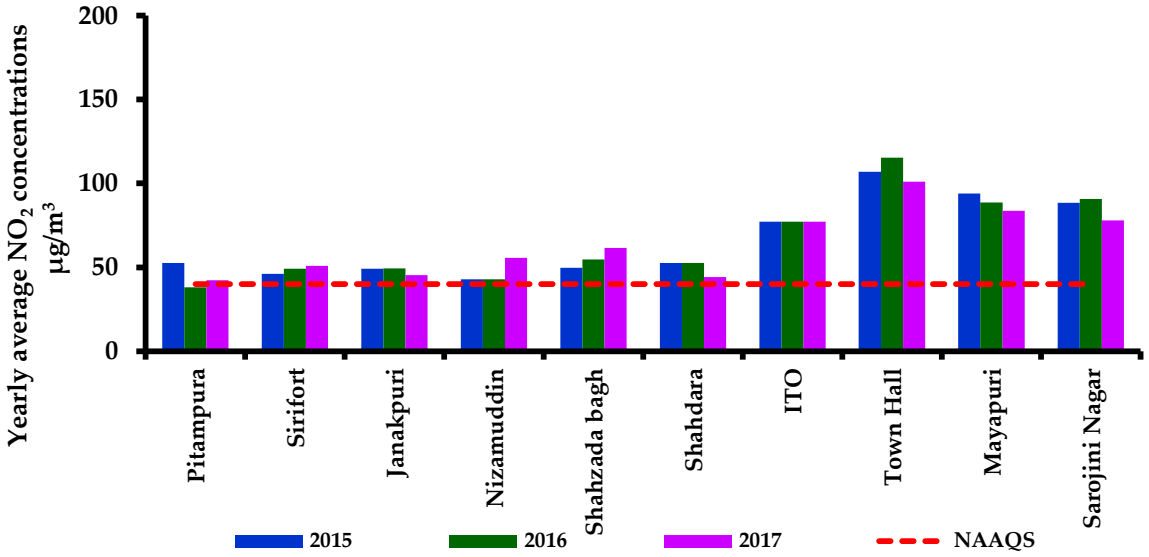
Ambient air quality trends have been analyzed for different stations in Delhi for two datasets – a) years 2015-2017 and b) 2018 (Jan-Dec) to assess the state of air quality. The data was plotted for yearly average concentrations for each category of pollutant for different stations in Delhi and was compared with the prescribed standards.

Data from Manual monitoring stations

The annual average concentrations of PM_{10}, $PM_{2.5}$ and NO_2 at different locations are plotted in Figure 11.

As shown in Figure 11, yearly average NO_2 concentrations are found to be exceeding the NAAQS at most of the stations. Among all the monitoring stations in Delhi, ITO, Mayapuri, Town Hall and Sarojini Nagar are showing higher annual average NOx levels, having concentrations exceeding more than two times the annual average limit probably because of the contribution from vehicles.

Yearly average PM_{10} concentrations have been exceeding the NAAQS of 60 µg/m^3 at all the station over the last three years with average concentration ranging from 214-266 µg/m^3. As in case of $PM_{2.5}$, the yearly average $PM_{2.5}$ concentrations have exceeded the NAAQS of 40 µg/m^3 at all the manual monitoring stations with average concentration ranging from 85-97 µg/m^3.



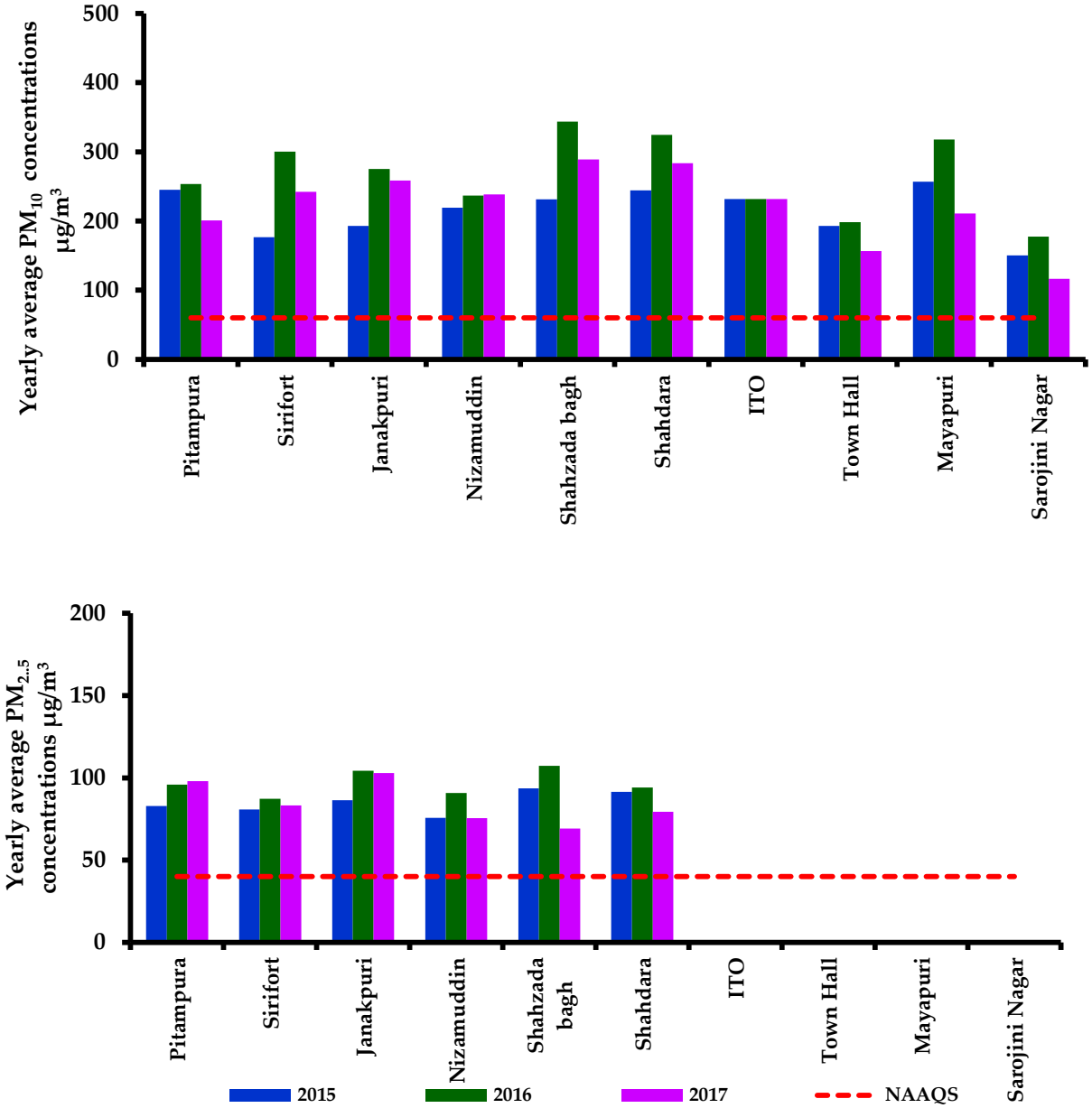


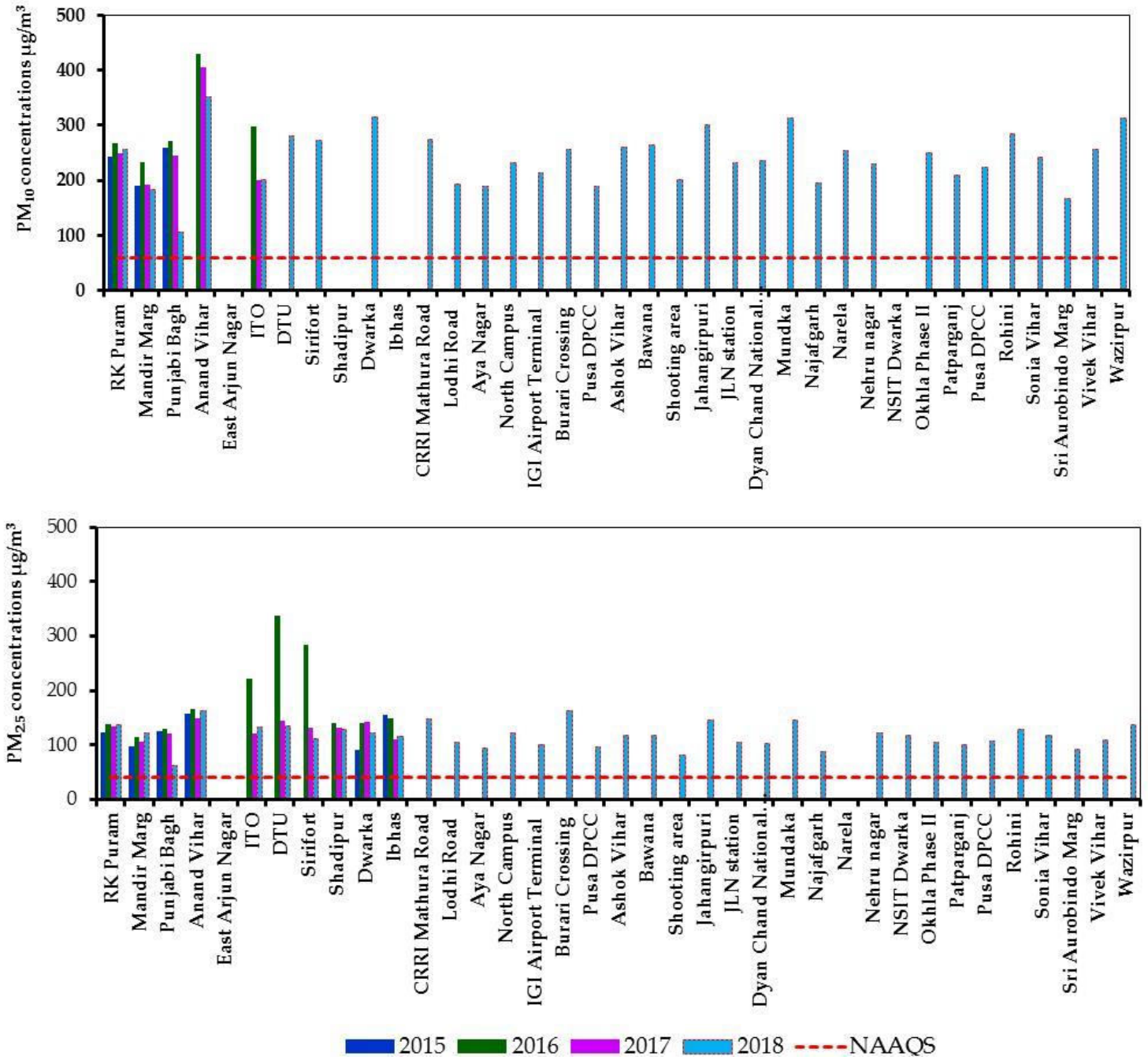
Figure 11 Yearly average concentrations of NO_2, PM_{10} and $PM_{2.5}$ for 10 manual air quality monitoring locations across Delhi

Data from continuous air monitoring stations

The annual average concentrations of PM_{10}, $PM_{2.5}$, O_3 and NO_2 at different continuous monitoring stations monitored in Delhi are shown in Figure 12. It is clear that the number of stations were drastically increased for the year 2018 compared to previous years.

It is evident from Figure 12 that NO_2 concentrations exceeded the standard at most of the continuous monitoring locations with average concentrations ranging from 35-70 µg/m^3. As in case of data from manual monitoring stations, the yearly average PM_{10} concentrations exceeded the standard at all stations with average concentrations ranging from 231-300 µg/m^3. In the last four years viz. 2015-2018, the lowest yearly average concentration was recorded in 2015 (231 µg/m$^{3)}$ which increased to 300 µg/m^3 in 2016. The yearly average concentration of $PM_{2.5}$ show that 2016 had the highest concentration of 182 µg/m^3 while the concentrations were comparatively lower in 2015, 2017 and 2018 with yearly average

concentrations of 125 and 129 and 117 μg/m³ respectively. However, the concentrations were still significantly higher than the standard value of 40 μg/m³. The yearly average ozone concentrations at different stations ranged from 37 -57 μg/m³.



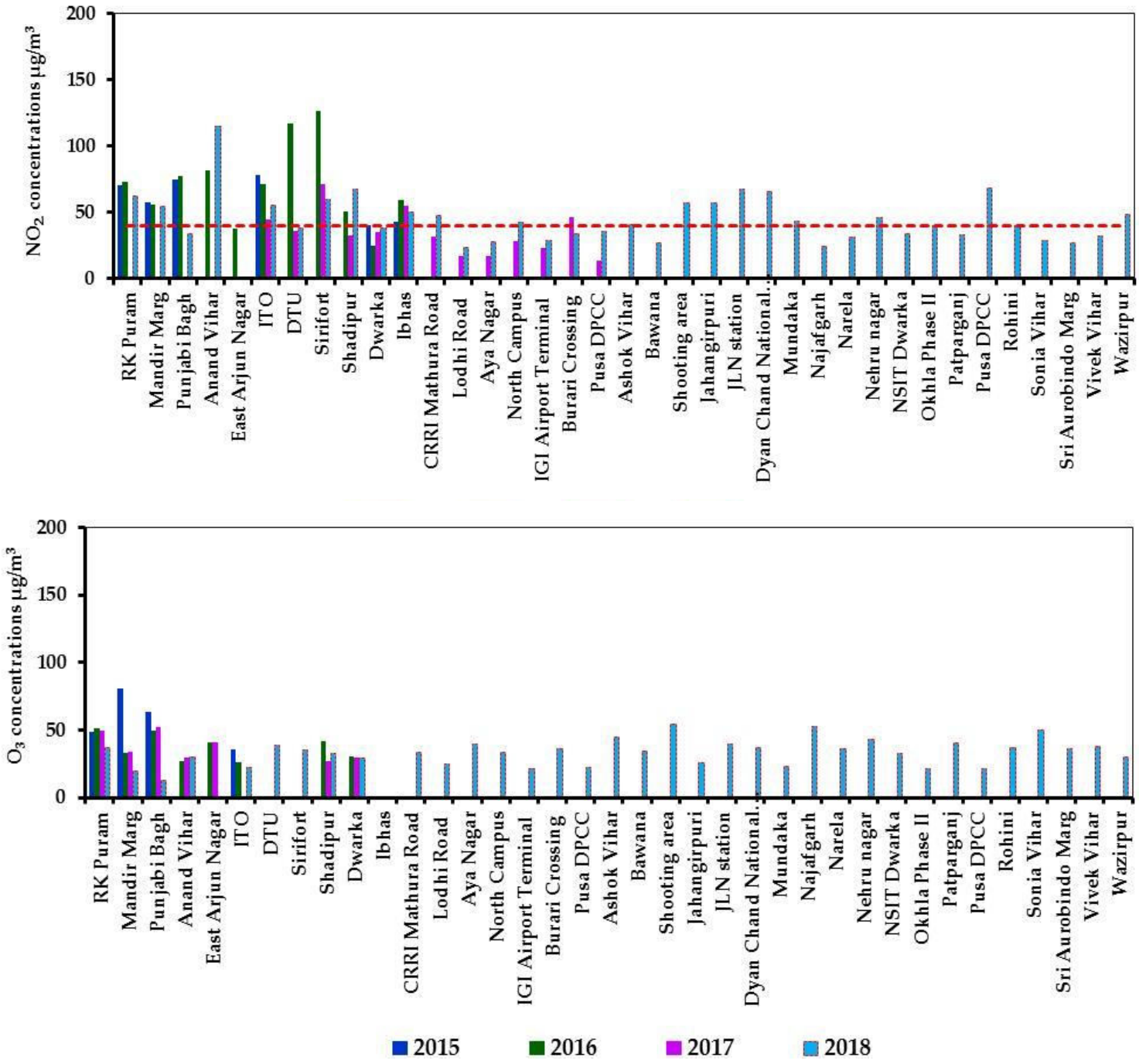


Figure 12 Yearly average concentrations (2015-2018) of PM_{10}, $PM_{2.5}$, NO_2, and O_3 for 37 continuous air quality monitoring locations across Delhi.

It is evident from Figure 12 that some data is missing for some of the continuous monitoring stations for different years (Annexure-I).

Ozone data has been specifically investigated for hourly values and has been presented in Figure 13. As an example, the values at Dwarka station were analysed, which show stark seasonal variation. The ozone concentrations are more in summers and autumn and less

during winters and rainy season. This is primary due to changes in patterns of solar radiations. Ozone is formed due to reactions of NOx and VOCs in presence of sunlight, and hence follows a seasonal pattern.

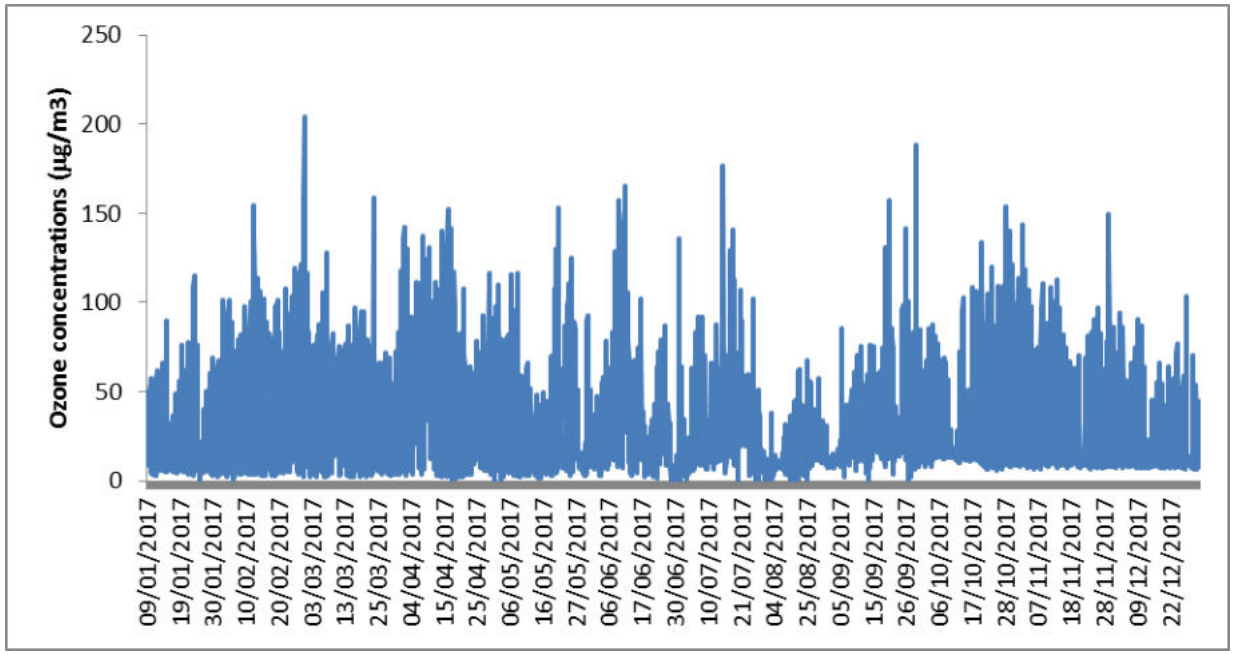
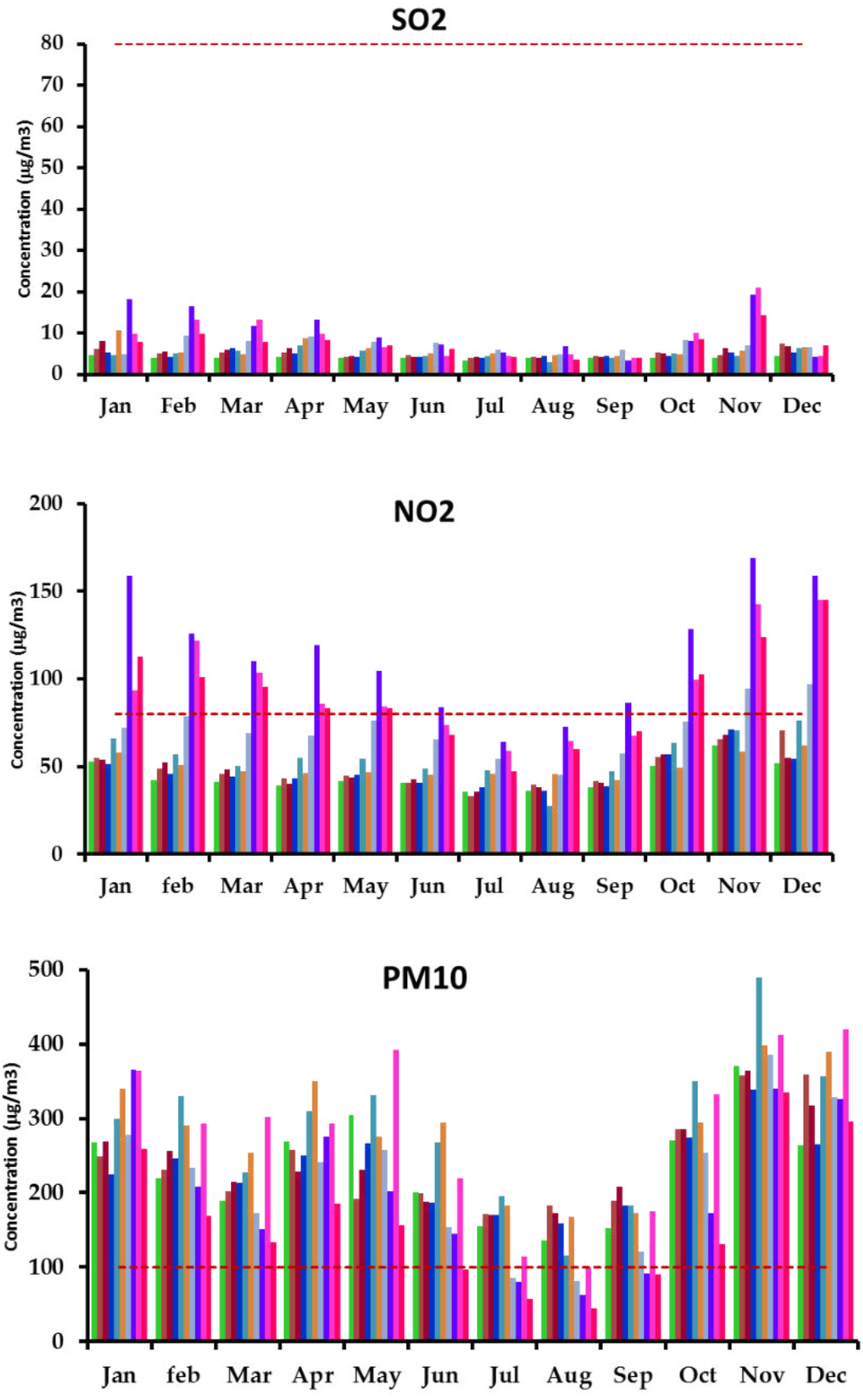


Figure 13 Hourly variation of ozone concentrations in 2017 at Dwarka monitoring station in Delhi

Seasonal variation

The data for different months in the years 2015-17 have been averaged to assess seasonal variation of pollutant concentrations. Figure 14 shows the three years averaged monthly concentrations for different manual CPCB stations for SO_2, NOx, PM_{10} and $PM_{2.5}$. Clearly, SO_2 concentrations are well within the limits. Mayapuri, Saronijini Nagar and Town hall show the highest NOx concentrations throughout the year. The values are highest during winters due to lower wind speeds and lower planetary boundary layer height. This reduces the dispersive capacity of the atmosphere. Rainy season shows the downwash effect and pollutant concentrations are reduced to the lowest levels in the year. PM concentrations also follow the same seasonal trends, however, the values are found to be highest in November, mainly due to additional emission contributions from agricultural burning in the neighbouring states.


SO2
Concentration (µg/m3)
80
70
60
50
40
30
20
10
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
NO2
Concentration (µg/m3)
200
150
100
50
0
Jan
feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
PM10
Concentration (µg/m3)
500
400
300
200
100
0
Jan
feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

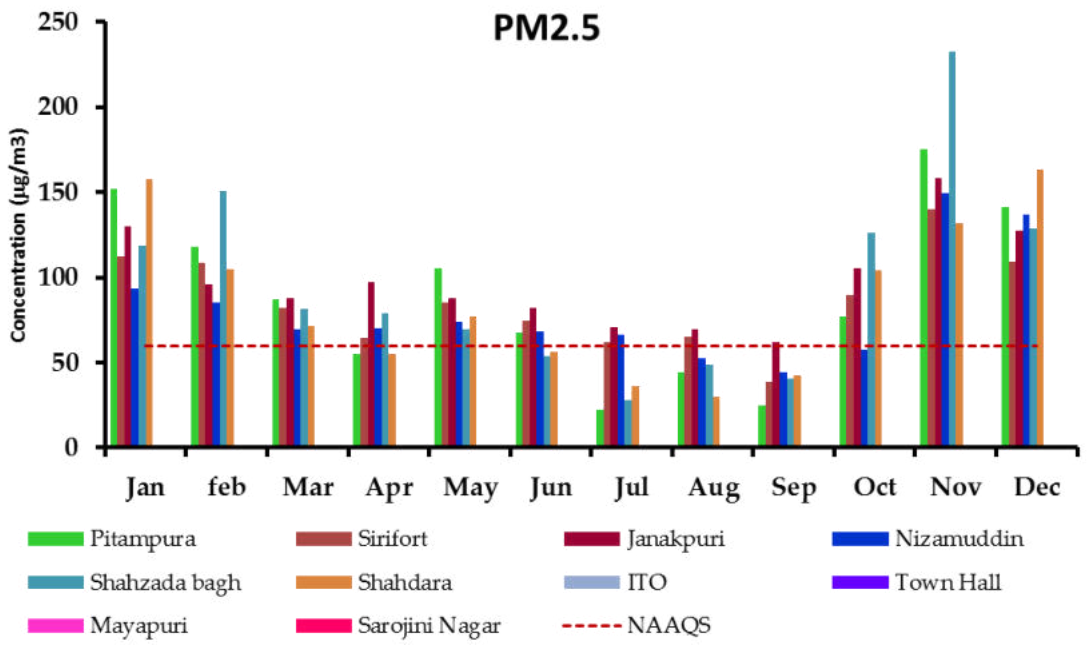


Figure 14 Seasonal variation of SO_2, NO_2, PM_{10} and $PM_{2.5}$ concentrations in different manual monitoring stations in Delhi (averaged for 2015-2017)

Ratio of concentrations of different pollutants

SO_2/NO_2

The ratio of SO_2 to NOx concentration has been analysed for various stations and months of the year. The ratio of SO_2 concentration to NO_2 concentration is almost constant through the years with the concentration of NO_2 remaining higher at all the locations suggesting significant contribution from vehicular sector (Figure 15). Shahdara, Mayapuri and Janakpuri stations showing the higher ratios, suggesting influence from industrial emissions.

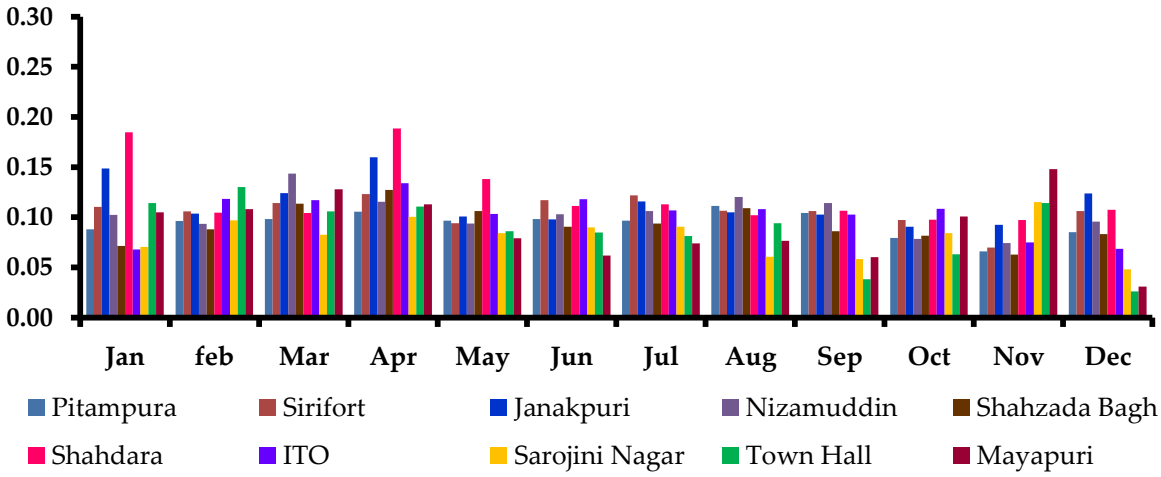


Figure 15 Monthly average SO_2 to NO2 ratio at different locations in Delhi (Averaged for 2015-2017)

NO_2/PM_{10}

NOx is primarily released by high temperature combustion in sources like vehicles and DG sets. Higher ratio of NO2 to PM suggests the influence of these sectors. The ratio of NO_2 concentration to PM_{10} concentration is highest at Sarojini Nagar, ITO, Town Hall and Mayapuri (Figure 16) suggesting the influence of vehicular activity. The ratio shows highest values during the monsoon period suggesting the downwash effect of PM, which leads to drastic settling of the particles.

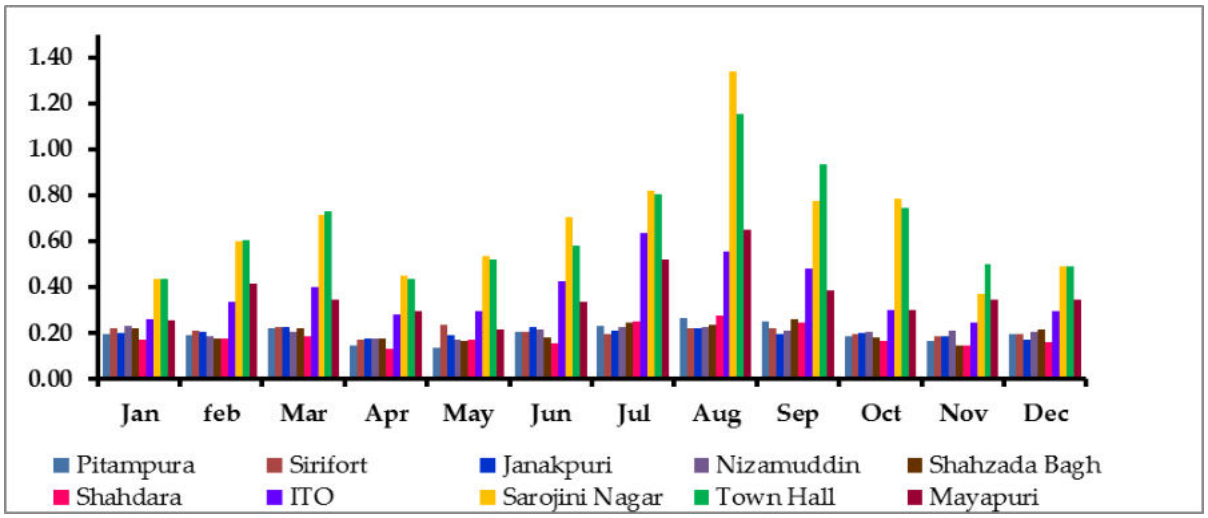


Figure 16 Monthly average NO_2 to PM_{10} ratio at different locations in Delhi (Averaged for 2015-2017)

$NO_2/PM_{2.5}$

The ratio of NO_2 concentration to $PM_{2.5}$ concentration is highest at Shahzada Bagh, Mayapuri and Pitampura. The values are highest during monsoon and post monsoon period (Figure 17) suggesting lowering of $PM_{2.5}$ concentrations due to rain downwash effect.

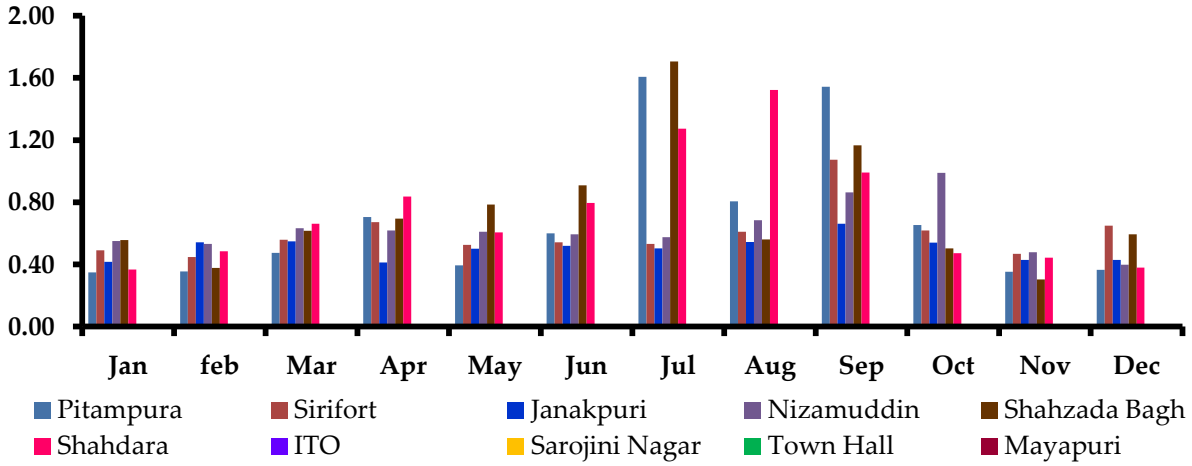


Figure 17 Monthly average NO_2 to $PM_{2.5}$ ratio at different locations in Delhi (Averaged for 2015-2017)

$PM_{2.5}/PM_{10}$

Combustion based sources show higher $PM_{2.5}$ fraction in PM_{10} emissions. On the other hand, dusty/crustal sources like road dust, construction, soil dust, flyash ponds, coal handling units etc show lower $PM_{2.5}$ to PM_{10} ratios. Higher ratio of $PM_{2.5}$ to PM_{10} shows the influence of combustion sources like transport, biomass burning, industries etc. The ratio of $PM_{2.5}$ concentration to PM_{10} concentration in Delhi is found to be higher during the winters compared to other seasons. This is due to increased influence of combustion based sources due to lesser dispersive atmospheric conditions in winters and lower wind speeds causing less re-suspension of dust.

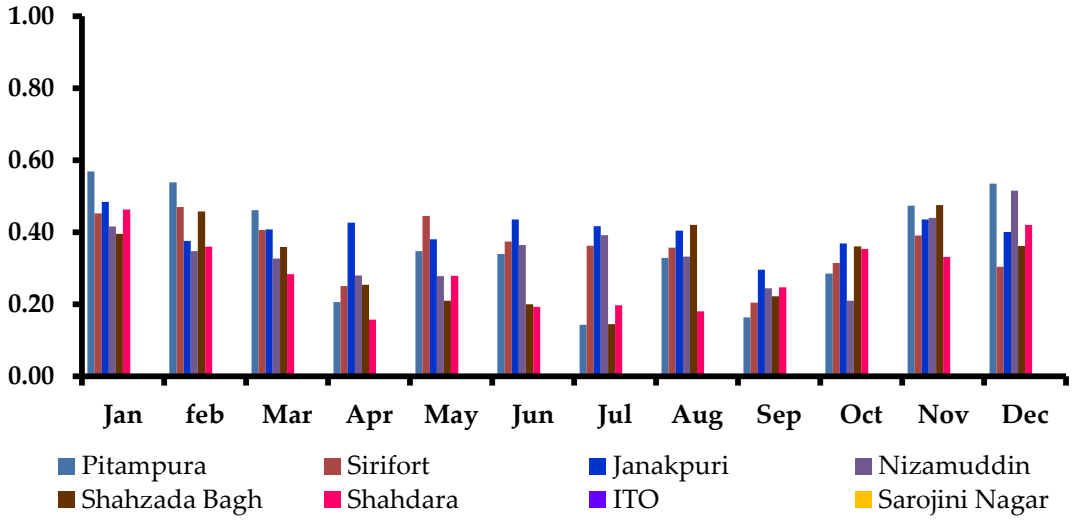


Figure 18 Monthly average PM2.5 to PM10 ratio at different locations in Delhi (Averaged for 2015-2017)

Correlation of pollutants among the manual monitoring stations

A correlation analysis has been carried out for different monitoring stations for various pollutants and is discussed in subsequent sections.

SO_2

In comparison to other pollutants, SO_2 is found to be somewhat less correlated among the monitoring stations, except for stations like Janak puri-Shahdara, Pitampura-Janak puri, Sirifort-Jankapuri which showed decent correlations. This suggests that the sources of SO_2 are not equally spread-out across the city and there are certain locations which show relatively higher SO_2 concentrations than other.

Table 1 Correlation analysis for SO_2

	Pitampura	*Sirifort*	*Janakpuri*	*Nizamuddin*	*Shahzada Bagh*	*Shahdara*	*ITO*	*Town Hall*	*Mayapuri*
Sirifort	0.72								
Janakpuri	0.79	0.74							
Nizamuddin	0.48	0.48	0.70						
Shahzada Bagh	0.26	0.55	0.49	0.40					
Shahdara	0.69	0.47	0.78	0.30	0.42				
ITO	-0.11	0.06	0.03	0.07	0.63	-0.12			
Town Hall	0.42	0.07	0.63	0.40	0.12	0.50	0.26		
Mayapuri	0.19	0.01	0.47	0.52	0.12	0.10	0.38	0.85	
Sarojini Nagar	0.31	0.20	0.56	0.42	0.31	0.21	0.48	0.80	0.92

NO_2

NO_2 concentrations in Delhi seem to be highly correlated across different monitoring stations. The correlation values vary between 0.78-0.96. The high correlation values suggest that the sources of NOx are widespread across the city. This can be attributable to vehicular emissions, which are widespread all across the city.

Table 2 Correlation analysis for NO_2

	Pitampura	*Sirifort*	*Janakpuri*	*Nizamuddin*	*Shahzada Bagh*	*Shahdara*	*ITO*	*Town Hall*	*Mayapuri*
Sirifort	0.90								
Janakpuri	0.95	0.88							
Nizamuddin	0.97	0.86	0.95						
Shahzada Bagh	0.83	0.85	0.79	0.81					
Shahdara	0.85	0.91	0.79	0.77	0.81				
ITO	0.82	0.89	0.84	0.84	0.92	0.80			
Town Hall	0.91	0.92	0.89	0.87	0.88	0.88	0.87		
Mayapuri	0.78	0.91	0.86	0.81	0.80	0.83	0.93	0.87	
Sarojini Nagar	0.84	0.97	0.86	0.80	0.87	0.89	0.91	0.95	0.93

PM_{10}

PM_{10} also shows high correlation among most of the monitoring stations. The values ranged between 0.74-0.96. This again suggests the influence of similar sources to PM_{10} concentrations across the stations. This also indirectly points out to significant influence of background contributions from outside of the city, which remain almost consistent for different stations in Delhi.

Table 3 Correlation analysis for PM_{10}

	Pitampura	*Sirifort*	*Janakpuri*	*Nizamuddin*	*Shahzada Bagh*	*Shahdara*	*ITO*	*Town Hall*	*Mayapuri*
Sirifort	0.74								
Janakpuri	0.84	0.93							
Nizamuddin	0.95	0.81	0.91						
Shahzada Bagh	0.94	0.81	0.90	0.96					
Shahdara	0.86	0.85	0.83	0.82	0.88				
ITO	0.94	0.87	0.95	0.94	0.94	0.93			
Town Hall	0.82	0.78	0.81	0.74	0.79	0.93	0.91		
Mayapuri	0.89	0.74	0.84	0.87	0.86	0.86	0.94	0.85	
Sarojini Nagar	0.83	0.87	0.91	0.82	0.84	0.91	0.95	0.95	0.86

$PM_{2.5}$

The concentration of $PM_{2.5}$ also suggests higher correlation having significant value reaching up to 0.95. This suggests the influence of similar sources to $PM_{2.5}$ concentrations across the stations. Other than vehicular contributions which are widespread across the city, this also indirectly points out to significant influence of background contributions from outside of the city, which remain almost consistent for different stations in Delhi.

Table 4 Correlation analysis for $PM_{2.5}$

	Pitampura	*Sirifort*	*Janakpuri*	*Nizamuddin*	*Shahzada Bagh*
Sirifort	0.96				
Janakpuri	0.91	0.91			
Nizamuddin	0.86	0.86	0.89		
Shahzada Bagh	0.86	0.91	0.90	0.80	
Shahdara	0.90	0.86	0.87	0.79	0.77
ITO					
Town Hall					
Mayapuri					
Sarojini Nagar					

4. Spatial mapping model for Delhi

The model was setup for two distinct air quality datasets – 2015-2017 and 2018. The primary reason for using two datasets is because of inclusion of several new monitoring stations in the city of Delhi in the year 2018, which could lead to more reliable model setups. Section 4.1 shows the spatial model result based on air quality data set for limited number of stations during 2015-2017. On the other hand, section 4.2 shows the results of the spatial model based on the air quality datasets for many more monitoring stations although for the small period of January-2018 to December 2018.

Before running the spatial model for generation of maps of air quality, the data was filtered on account of QA/QC procedures. Annexure- II shows the procedure used for data filtering.

The data has been cleaned up for both the time periods 2015-2017 and 2018 (shown in Annexure II) and the summary of the stations used in RIO 2018 is shown in table below

Table 5 Summary of stations used in RIO 2018

Pollutant	No. of stations having		
	Limited data	Insufficient data quality	Used
PM_{10}	5	Nil	27
$PM_{2.5}$	4	1	30
O_3	4	3	28
NO_2	4	6	26

4.1 Spatial model results based on air quality data 2015 to 2017

After the clean-up of the data, the monitoring station long term averages were calculated for the whole period 2015-2017. To start with, a correlation analysis has been performed to understand the relationships between several variable with the air quality values observed at different monitoring stations. From the given latitude and longitude values, we have calculated UTM coordinates (X,Y) (projection system given in Annexure). Information about altitude of these stations was obtained from the GTOPO30 global terrain dataset (https://lta.cr.usgs.gov/GTOPO30). The station types were broadly defined in the following categories:

- Rural Area
- Residential Area
- Airport location
- Traffic
- Industrial

The station codes were defined as short tags based upon the full station names, see annex. Table 6 shows the station codes, UTM coordinates, altitude, and average concentrations of PM_{10}, $PM_{2.5}$, NO_2, and O_3.

Table 6 Details of monitoring stations and their ambient air pollutant concentrations (annually averaged)

ID	Station name	STATCODE	X [m]	Y [m]	ALT [m]	TYPE	PM_{10}	$PM_{2.5}$	NO_2	O_3
1	Mandir Marg	MANDIR	128578.78	3173551.81	229	1	199.1	104.4	55.1	
2	Siri Fort	SIRIFO	129327.94	3163953.64	217	2		130.6	79.9	
3	Delhi Tech. University	DTU111	120750.28	3186519.4	212	1		157.4	73.0	
4	RK Puram	R1K1PU	126903.02	3165470.02	219	2	234.9	127.0	69.0	48.2
5	IGI airport	IGI1AI	117188.33	3164903.92	224	3		82.0	55.6	86.3
6	Dwarka	DWARKA	112221.98	3171114.36	210	2		120.6	33.6	30.7
7	Lodhi road	LODHI1	130425.17	3168362.2	208	2			59.7	
8	ITO	ITO111	132430.16	3172599.06	206	4	220.7	138.1	60.7	36.8
9	Anand Vihar	ANAND1	139780.13	3174388.02	199	5	279.1	152.7	84.2	28.3
10	Civil Lines	CIVIL1	130996.72	3177974.83	209	1				79.4
11	East Arjun Nagar	EAST1A	137798.32	3175421.71	198	5			33.6	41.1
12	Shadipur	SHADIP	124002.84	3175775.01	216	4		141.0	52.8	31.9
13	IHBAS hospital	IHBAS1	138889.19	3178409.42	201	2		110.7	51.7	

As the RIO geospatial mapping model is based on the correlation of the long term concentration values with a number of well-chosen proxy parameters, the correlation analysis was carried out to understand these relationships.

4.1.1 Spatial correlation analysis 2015-2017

When enough station data is available which can be considered representative for different land cover/land use categories, the preferred way to deploy the RIO model is making use of a land cover dataset as described in (Janssen et al., 2008). However, in this case the number of stations is limited with valid data and from these limited number of stations, especially for PM_{10} (4 stations), it is not expected to fit the same number of degrees of freedom as in the setup elaborated in (Janssen et al., 2008). The model takes a minimum of five station data to interpolate the spatial map. If in case we use four stations, the areas far off from the stations will be plotted with the available data set and this reduces the degree of accuracy/freedom in plotting the maps.

In this correlation analysis we have considered:

a) The Global Human Settlement Layer (GHS - http://ghsl.jrc.ec.europa.eu/), which is a population dataset for 2015, disaggregated to 250 m resolution using satellite imagery (Figure 19). The natural logarithm (LN(1+P)) of the mean population was used in the buffer area to bring the variation in population density more into accordance with the variation in local air quality. This had been shown to improve the correlation in many other regions

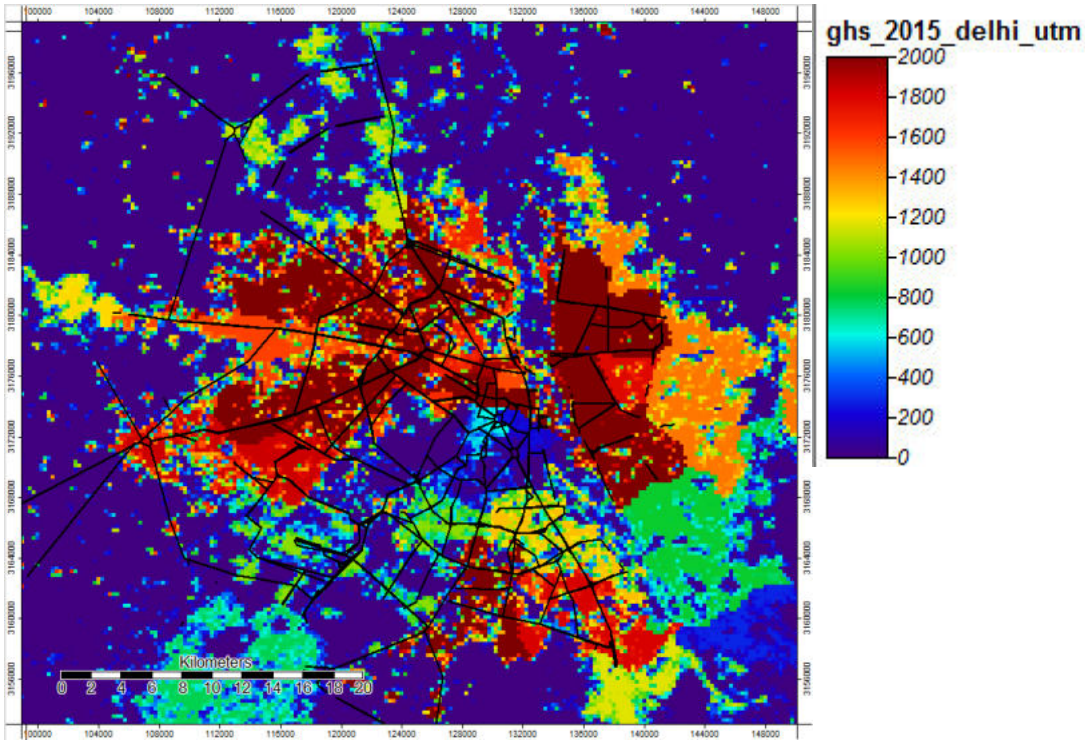


Figure 19 Visualization of the GHS dataset for Delhi - 2015 gridded at 250m.

b) The road network in 3 different categories: major roads, minor roads and sub arterial roads, which has been digitized by TERI. The road lengths were aggregated for the different road categories in the different buffer sizes (see further). In the different buffers, the normalized total road lengths w.r.t. the buffer area have been used to make to values independent of the buffer area.

c) The LULC dataset (Google earth) has been used which contains the different land use land cover types as depicted below in Figure 20 .

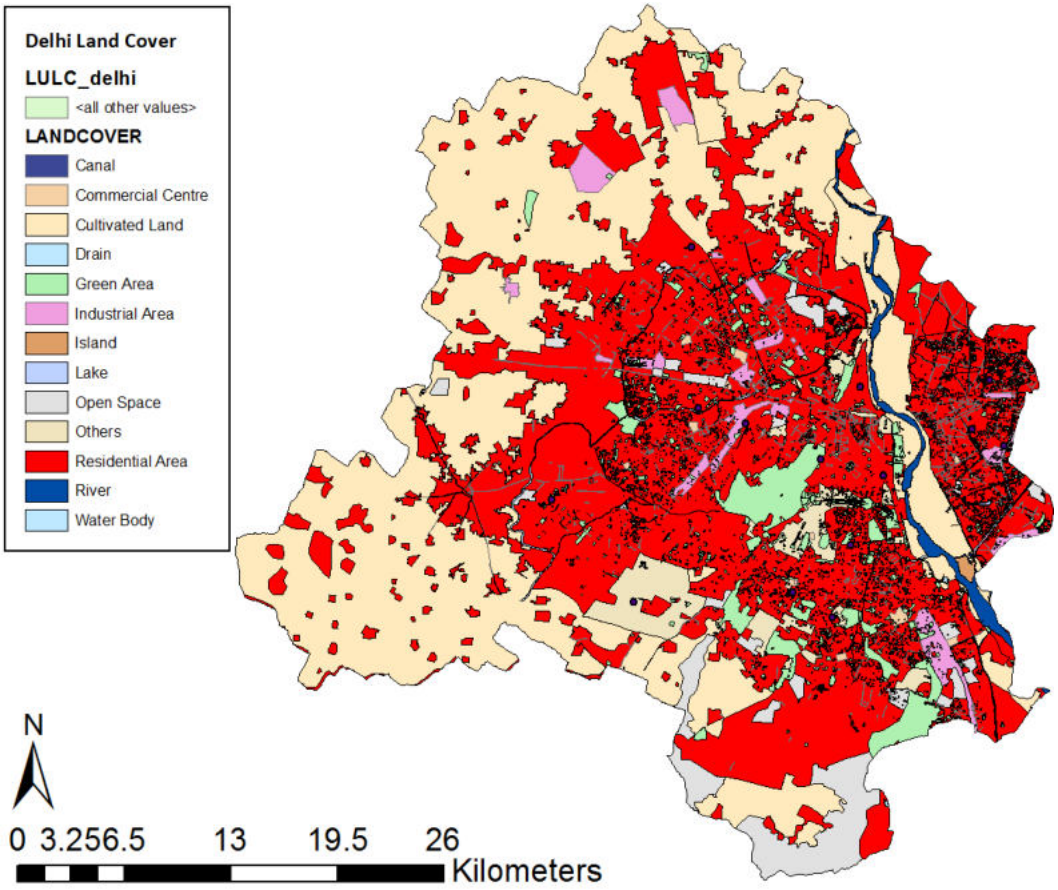


Figure 20 Land cover dataset used for the analysis.
Source : Google Earth

Also, here for each station we have created **different buffer sizes** (500 m, 1km, 1.5 km, 2 and 3km) to calculate the total fraction of the different land cover classes covering the buffer around the stations. Note that for each LULC class, the fraction within the buffer area has been used.

d) Finally, we also investigated the correlation with altitude using the GTOPO30 dataset to derive the station altitude values.

As mentioned above here, we have calculated the proxy parameters in different station buffers (circular of 500 – 1000 – 1500 - 2000 and 3000 m radius) to assess the dependency on spatial scale of the observed spatial variation in concentration. The key features of the correlation analysis and the conclusions for the deployment of the RIO model are presented below. The correlation analysis also reveals the most appropriate variable for carrying out spatial modelling for different pollutants.

PM_{10}

In case of PM_{10}, the correlation analysis is somewhat difficult given the limited number of stations (4). The observations are listed below:

- There seems to be consistently over the different buffer radii tested a negative correlation with altitude (r = -0.8) and a positive correlation with the population density which is the strongest (r=0.9) at small buffer size (500m - 1 km). We find also a high correlation with Industrial land cover (however, this appears to be

caused by only a single station (and is therefore not really full-proof to be used in this analysis).

- The 2015-2017 data seem to exhibit a negative correlation with the major and sub arterial road proxies and a slightly positive correlation with the minor roads (which may be more the case given the correlation with residential fraction for the minor roads). In any case, the negative correlation with the major & sub arterial roads seems a bit odd but given the limited amount of stations, we should not over interpret these findings. Nevertheless, we did expect a stronger correlation given the contribution of road dust emissions in the coarse PM fraction.
- In summary, considering the high correlations with PM_{10} data, it seems reasonable to setup a PM_{10} mapping model only using the GHS_2015_MEAN data proxy, for which the 500 m – 1 km buffers seem to correlate best (Figure 21).

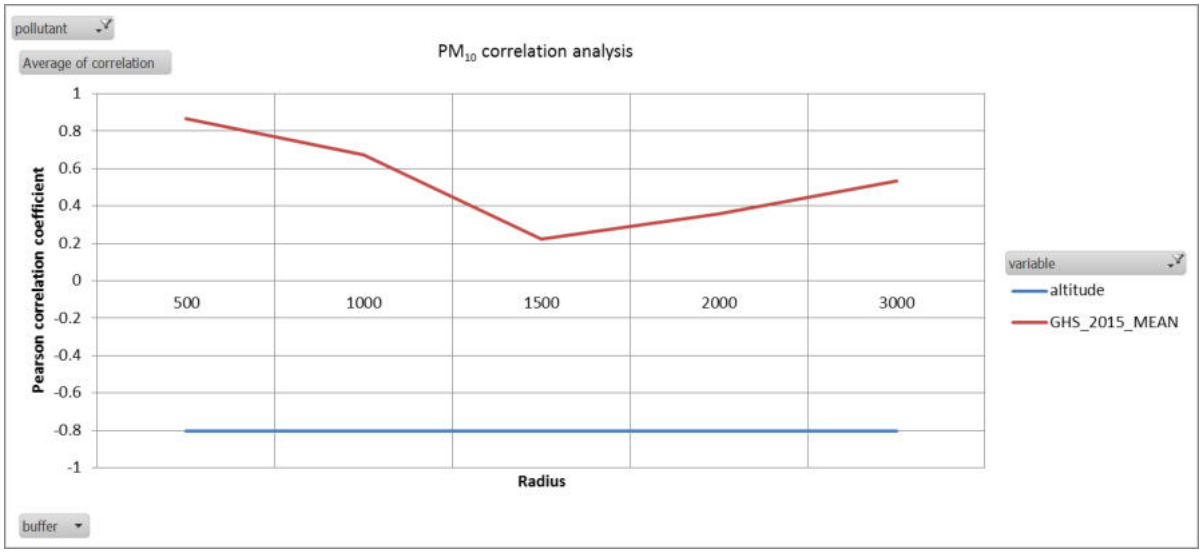


Figure 21 Coefficient of correlation (and its variation with the buffer radii) with GHS_2015_MEAN population data and altitude with PM_{10} values at different stations.

$PM_{2.5}$

In case of $PM_{2.5}$, the correlation analysis is somewhat better given more number of stations reporting valid data for $PM_{2.5}$. The observations are listed below:

- Residential land cover class shows quite a strong correlation across the different buffer radii, but the GHS population dataset does not correlate so strongly with $PM_{2.5}$ concentrations across different stations (Figure 22).
- Like PM_{10}, significant negative correlation of $PM_{2.5}$ concentrations with altitude seems to be present.
- Next to residential area, the industrial and "Green Area" area fractions are also consistently correlated across the scales, though for Green Area, we see the (negative) correlation diminishing with increasing spatial scale (Figure 23).

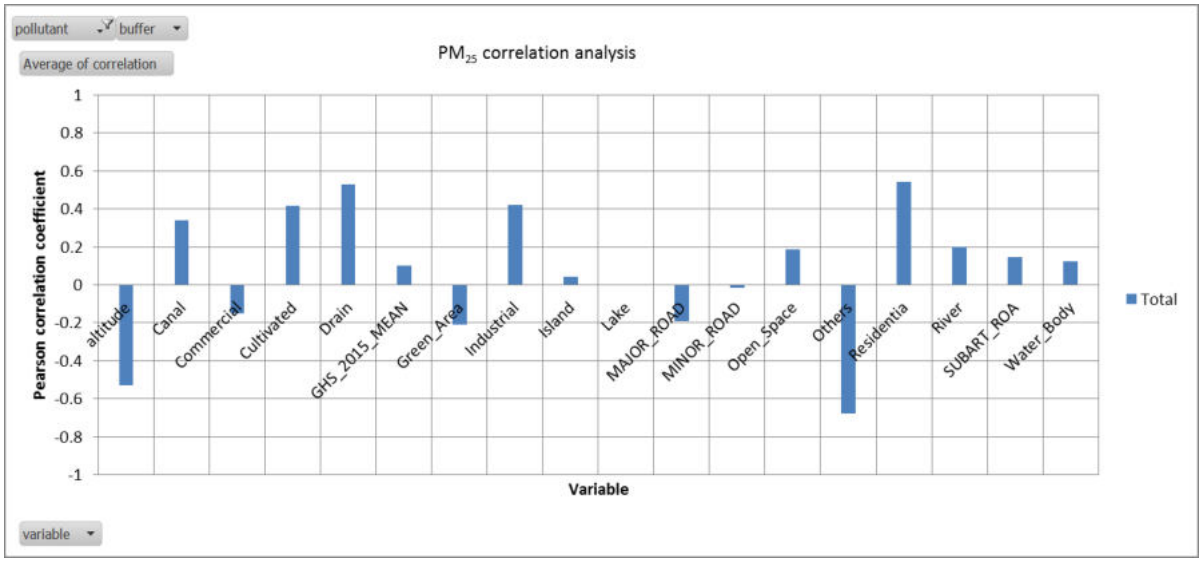


Figure 22 Coefficient of correlation of $PM_{2.5}$ concentrations with different landuse categories at different stations (averaged for different buffer sizes)

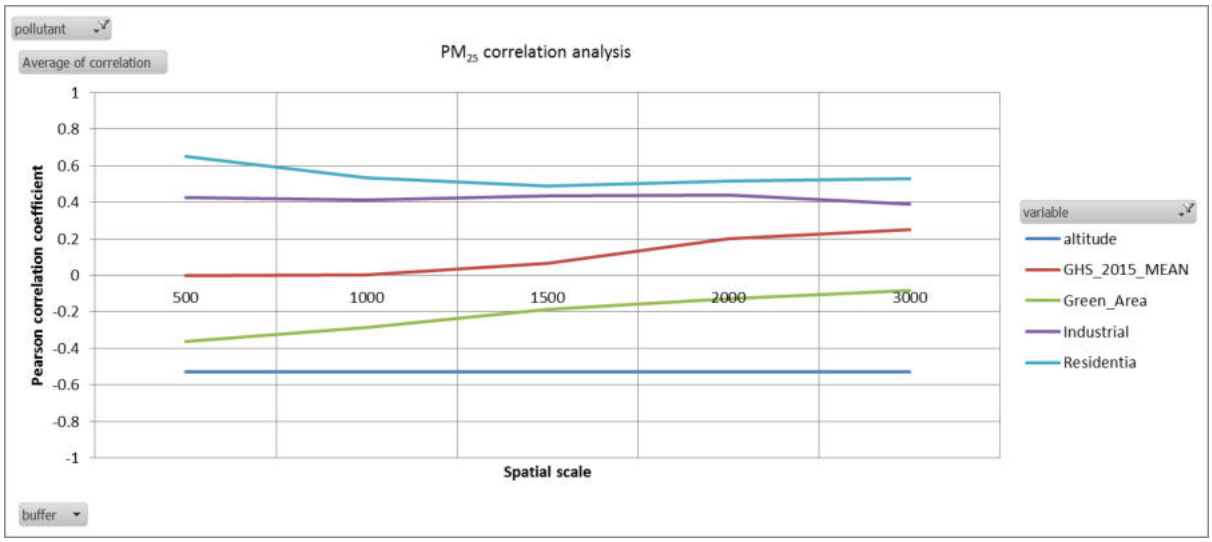


Figure 23 Coefficient of correlation (and its variation with the buffer radii) with GHS_2015_MEAN population, altitude, and landuse categories with $PM_{2.5}$ values at different stations.

Ozone

In case of ozone, the correlation analysis is somewhat better given more number of stations reporting valid data for ozone. The observations are listed below:

- Consistent correlation with altitude seems to be present (positive, i.e. higher O_3 with increasing altitude due to stratospheric influx) (Figure 24).
- Stable correlation across the scales for population density and / or residential surface fraction, which seems to exhibit strongest correlation at spatial scales of about ~2 km buffer radius (Figure 25). These correlations are negative clearly depicting the titration chemistry of ozone with NO emissions released in the residential and other dense areas.

- Sub arterial road fraction seems to be consistently negatively correlated which again makes sense due to titration effect, however the positive correlation of the total major road length seems to be odd, although this only occurs at small spatial scale (500), at larger scale, there is hardly any correlation, see below.

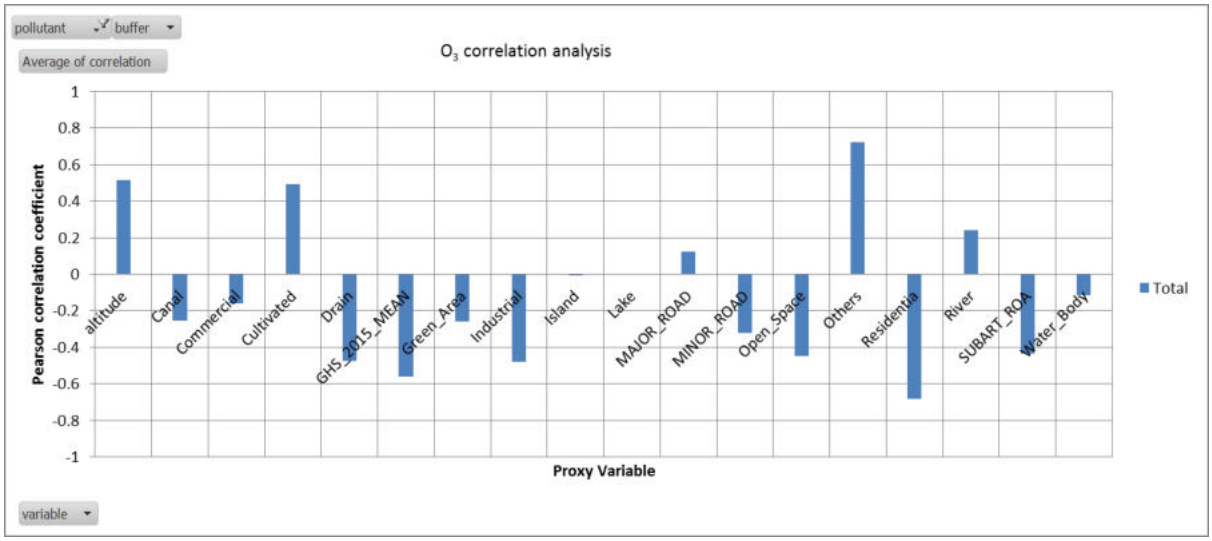


Figure 24 Coefficient of correlation of ozone concentrations with different landuse categories at different stations (averaged for different buffer sizes)

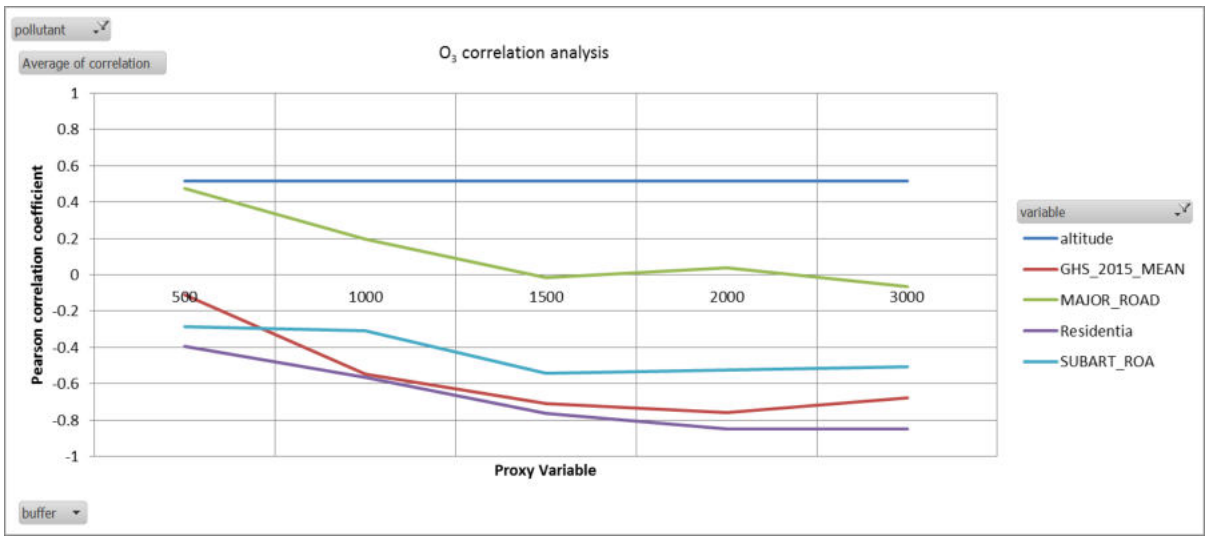


Figure 25 Coefficient of correlation (and its variation with the buffer radii) with GHS_2015_MEAN population, altitude, and landuse categories with ozone values at different stations.

NO_2

In case of NO_2 (Figure 26 and 27), the correlation analysis is in general found to be unsatisfactory, even for population density, residential area fraction, road length etc. The correlation with the most relevant proxy parameters is found to be negative across the different spatial scales. Some further detailed analysis is needed before being able to proceed to construct an NO_2 mapping model. This could include traffic count information (annual averaged daily traffic counts) or including a dispersion model, or an emission dispersion based proxy parameters.

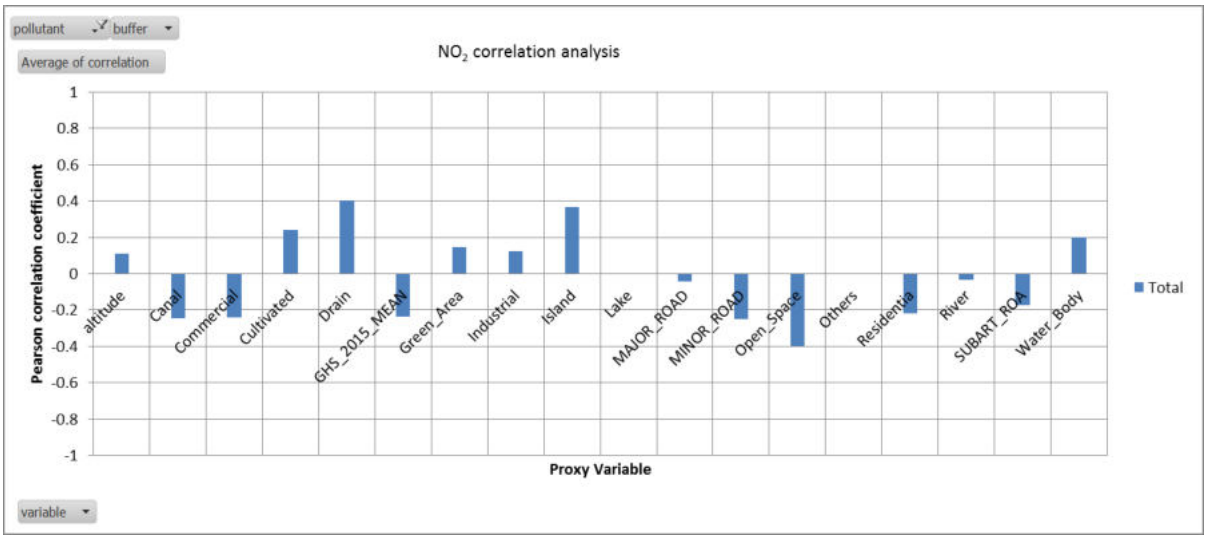


Figure 26 Coefficient of correlation of NO_2 concentrations with different landuse categories at different stations (averaged for different buffer sizes)

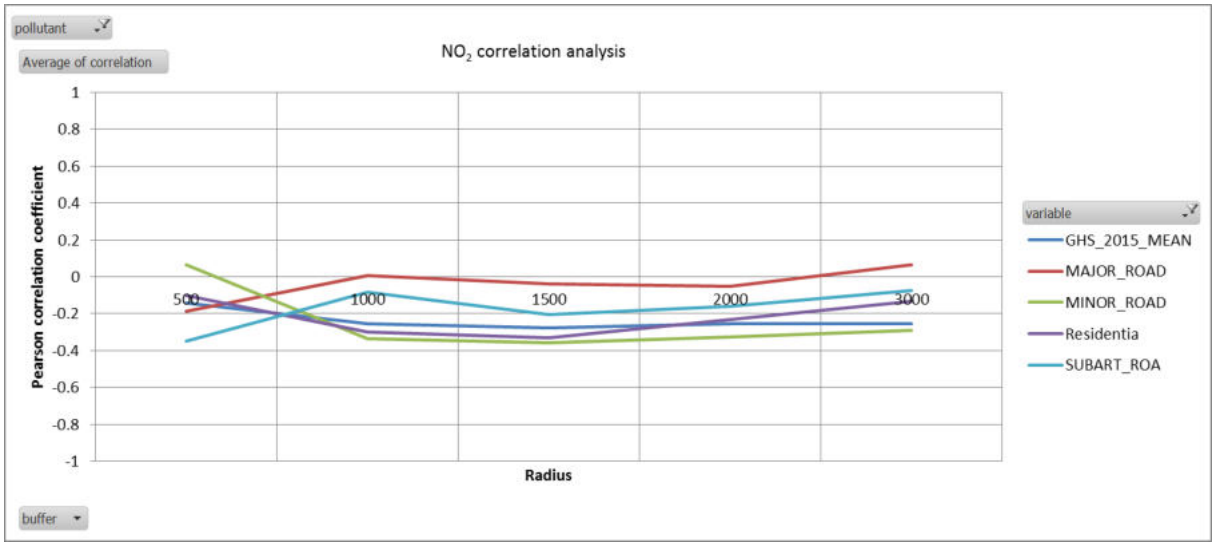


Figure 27 Coefficient of correlation (and its variation with the buffer radii) with GHS_2015_MEAN population, roads, and landuse categories with NO2 values at different stations.

The correlation analysis suggests that there are variable which are found to be significantly correlated with the concentrations of PM_{10}, $PM_{2.5}$ and O_3. However NO_2 requires some further study as none of the variables were found to be suitably correlated with the

concentrations. In this phase, we then developed spatial models for PM_{10}, $PM_{2.5}$ and O_3 concentrations only.

The simple first step is to take the population density map as a spatial driver for all, as this seems to yield reasonable correlations and is intuitively related to sources leading to air pollution. We therefore go ahead with the RIO – optimization technique and constructed a reduced set of land use categories based upon the LULC dataset. Limited number of stations is a real constraint in this exercise for the years 2015-2017.

4.1.2 Construction of RIO spatial mapping model (based on 2015-2017 dataset)

RIO model is constructed for the city of Delhi to produce pollutant concentration estimates on a 1x1 km2 grid over the city. The basics of the RIO model are described in detail in Hooyberghs et al. (2006), and Janssen et al (2008). The first step is to develop a grid of 1x1 km resolution grid across the city of Delhi and overlay the landuse map on it. Locations of all the monitoring stations are marked on the study domain and a buffer zone dimensioned by a fixed radius is created around each of the monitoring station. Within the buffer zone, area under different landuse classes is identified and quantified. The next step is to establish β parameter, which is a single value indicator correlating the local land use characteristics with the air pollution levels originated due to local activities.

Establishing land use β parameter

In order to establish land use β parameter, we identified % area under different land use categories within buffer region around all monitoring stations in study domain. The defined single value indicator is based on correlation of local land use characteristics to local air pollution levels in which land cover distribution in the vicinity of all monitoring stations is transformed into single land use parameter β according to formula:

$$\beta = \log\left[1 + \frac{\sum_i a_i \times n_{\mathrm{RCL}i}}{\sum_i n_{\mathrm{RCL}i}}\right].$$

Where,

nRCLi is the area of landuse class i inside the buffer and ai is a pollution-related coefficient for the landuse class i. The β -parameter is the logarithm of a weighted and normalised sum of the landuse class distribution. The pollution-related coefficients ai are used to weigh the importance of a particular landuse class on the air pollutant concentrations. The β indicator can then be optimized for each pollutant by choosing a best set of ai coefficients, i.e which gives the best results. The different landuse classes used in this analysis are shown below in Table 7.

Table 7 RIO classes (based on land cover)

S.No.	Landuse class
1	a1 = Residential, Drain
2	a2 = Industrial, Commercial
3	a3 = Green Area
4	a4 = Cultivated
5	a5 = Open Space, Others
6	a6 = River, Water body, Canal, Lake, Island

For initial estimates of β, initial pollutant coefficient a_i is derived from source emission estimates for Delhi (Table 8). Emissions are used as the initial estimates for ai, as they provide the initial guess for the local characteristics of air quality. Sectoral emissions for Delhi NCT have been used from Guttikunda and Calori (2013). The estimate of ai values is shown below in Table 8.

Table 8 Initial estimates of ai derived from available sectoral emission inventory estimates

a	Landuse	Sectors for initial weights using emissions	Emission share (PM 2.5)	Scaled based on area of each landuse
a1	Residential area	S1+S2+S6+S8	0.4	1
a2	Industrial + Commercial	S3+S5+S9	0.36	0.9
a3	Urban Green	-	-	-
a4	Cultivated land	S7+S4	0.23	0.575
a5	Open space	-		0
a6	Water bodies	-		0

Where,

Sector	Label
S1	Transport
S2	Domestic
S3	Diesel Gen
S4	Brik Kilns
S5	Industries
S6	Construction
S7	Waste Burning
S8	Road Dust
S9	Power Plant

Thereafter, β -parameter is estimated for all the stations using the initial conditions for ai based on emission estimates, and area of different landuse categories in the buffer. The correlation between β values and pollutant concentration at different stations is observed and further optimized by validating the values using hit and trial method. Once optimized, the final set of ai and β values for each station are estimated.

Optimization of land use β indicator

Once the stations are characterised by a β value, the a_i values need to be optimised in order to improve the correlation between β and pollutant concentrations. Therefore, the long-term annual average of the pollutant value in each monitoring station is plotted against the corresponding estimated β value of the station. This plot represents ''trend'' function: it is the correlation between the mean pollutant value and the land use β -parameter for the near vicinity. The high correlation between average pollutant concentration and β –parameter can be understood as optimal parameterization of local land character of spatial distribution of air pollutant. The optimization is achieved by improving the coefficient of determination (r^2) and reducing the root mean square error (RMSE) of the fit and choosing the final set of a_i values. Thus, the optimization produces a statistically valid trend function between mean pollutant value and the land use β –parameter. The optimised ai values are presented in Table 9.

Table 9 Optimized ai values for different landuse categories

a		Values before optimization	Values after optimization
		PM2.5	PM2.5
1	Residential area	1	1.000
2	Industrial + Commercial	0.9	1.217
3	Urban Green	0.0	0.000
4	Cultivated land	0.575	0.575
5	Open space	0	0.161
6	Water bodies	0	0

The optimisation of β parameter has been done separately for different pollutants namely $PM_{2.5}$, PM_{10} and Ozone.

Optimization of β parameter with $PM_{2.5}$

The optimization model runs to derive ai values for $PM_{2.5}$ are also depicted in Figure 28. Figure 27 shows the correlation between $PM_{2.5}$ concentration and β parameter by making adjustments in pollutant related coefficient ai during optimization. It was observed that the fit shows r^2 value of 0.62, and RMSE of 13.28. It was observed for high β values the concentration was also high at residential and traffic locations, depicting the role of local activities on deterioration of air quality.

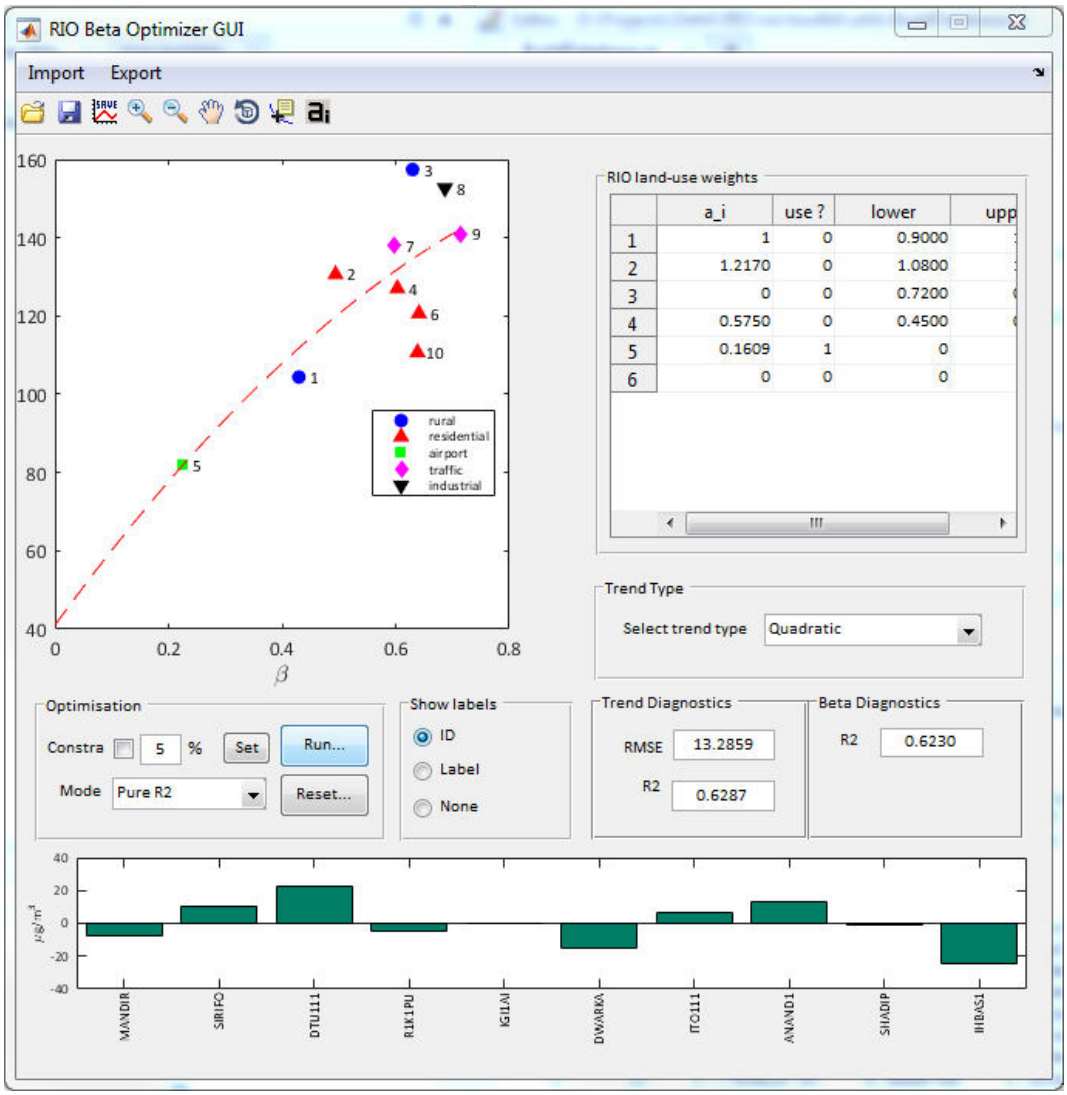


Figure 28 Optimisation model runs to derive ai values for $PM_{2.5}$

Optimization of β parameter with PM_{10}

The ai values have been assumed to be same for PM_{10} considering similarity of sources with $PM_{2.5}$ and limited number of PM_{10} stations (4). The trend fit of PM10 with β parameter yields a satisfactory correlation and trend function, as can be seen below in Figure 29.

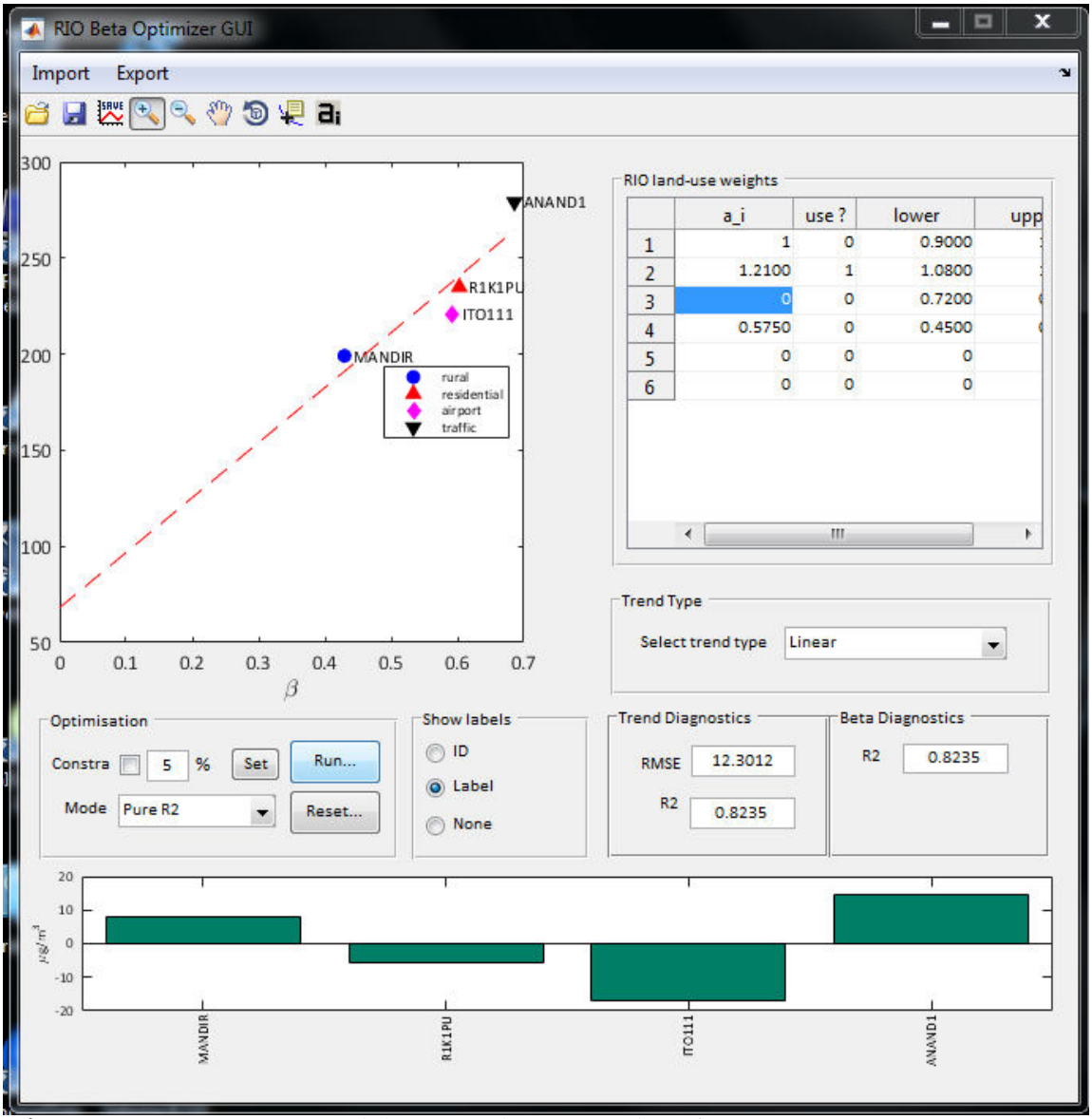


Figure 29 Optimisation model runs to derive ai values for PM_{10}

In case of PM_{10}, the optimization for establishing trend was performed using data for four available monitoring stations. The optimization yielded a linear plot showing high concentration at kerbisde , residential and airport bound stations corresponded with high β parameter with r^2 value of 0.82 and RMSE value of 12.30. From the trend analysis, highest average concentration reaching above 250 μg/m^3 was also observed at the station ANAND1, a traffic bound location indicating suspension of road dust along with vehicular emission exhaust could be major local contributor near this station.

Optimization of β parameter with O_3

In case of O_3, for optimization an interesting trendline was constructed, which yielded very high correlation, when allowing the weight for open space to vary as well. The set was obtained by first varying a2, a3 and a5 and keeping a4 fixed at 0.5, after which we optimized a4 with the rest fixed. The optimization produces plot showing cross relationship between ozone concentration and land use β parameter indicating low ozone concentration for high β values. The plot shows high ozone concentration at rural area and airport region having low

β values whereas low ozone concentration was observed at traffic bound, industrial and residential areas for high β values. This is mainly due to "NO-titration effect", which leads to depletion of ozone at the areas of high NO emissions (Sharma et al., 2016). NO being primarily released from vehicles, DG set (high temperature combustion activities) etc, is found to be more in the center of city and less on the periphery. Therefore, ozone is found to be higher in the periphery and less in the center of the city. Accordingly, it follows an inverse relationship with the β parameter (Figure 30).

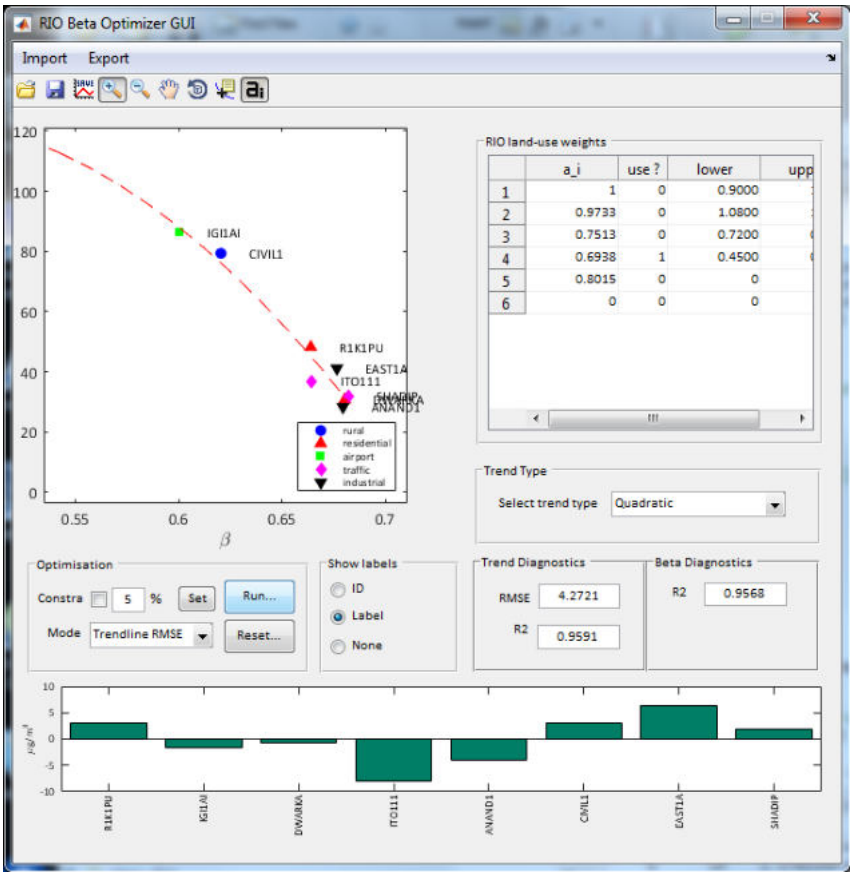


Figure 30 Optimisation model runs to derive ai values for ozone

4.1.3 RIO spatial mapping model validation 2017

The performance of the RIO interpolation model is evaluated using the "leaving-one-out" approach. In this approach, for each station concentrations are predicted making use of all other available monitoring data at that time step, except the concentrations of that particular station. For validation of model, different performance metrics such as – coefficient of correlation (r^2), root mean square error (RMSE), the bias and the mean absolute error (MAE) have been used. Also, to evaluate the overall performance of RIO model and also to make comparison with standard interpolation techniques, these metrics (RMSE, bias and MAE) have been used. These metrics have been computed for model performance for different pollutants namely, $PM_{2.5}$, PM_{10}, and ozone.

Model performance and validation: spatial scale

Validation of the model results for different pollutants is carried out by considering leaving one out approach. Such approach was carried out for all the stations in which model results

(after leaving out a particular station for estimation of β trends) were compared with the observed concentration value. It was observed from Figure 31 that modelled values for most stations are close to their observation in the year 2017; however, there are some stations which showed over and under estimation of predicted modelled values. Overall, the bias for $PM_{2.5}$ is -5.3 and RMSE of 26.9. In case of PM_{10} (Figure 31), there were only 4 stations for which analysis could be carried out which show a bias of 24 and RMSE of 72. It was observed from PM_{10} plot that optimized modeled results were close to their observation except one station showing over estimated value.

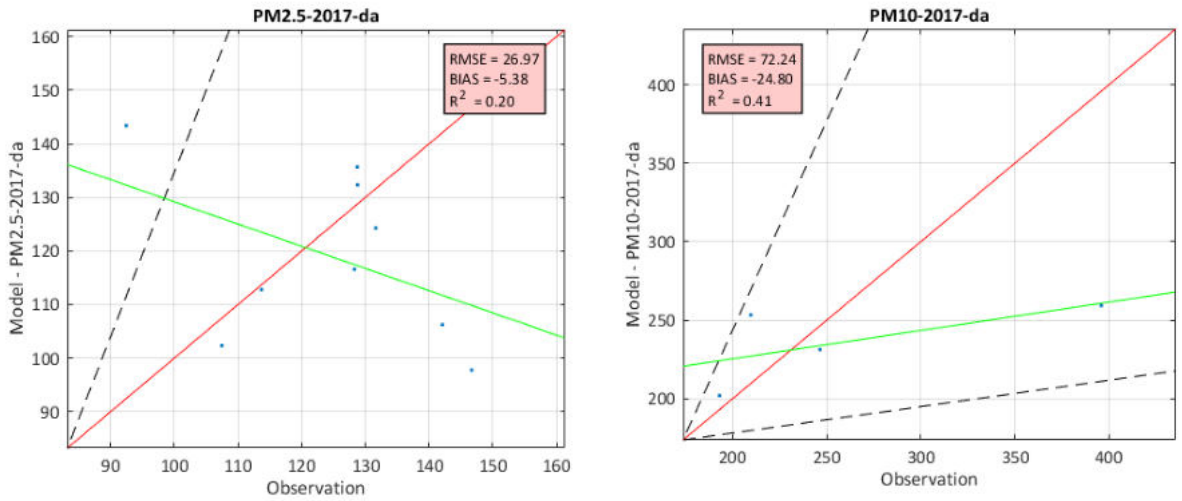


Figure 31 Comparison of modelled (using RIO) and actual PM2.5 and PM_{10} concentrations for 2017

In case of O_3, the analysis shows that the optimized modelled value for all stations were highly close to their observation having r2 value of 0.78, a very small bias of -0.69, and RMSE of 4.51. Out of the three pollutants, the model has performed the best for ozone.

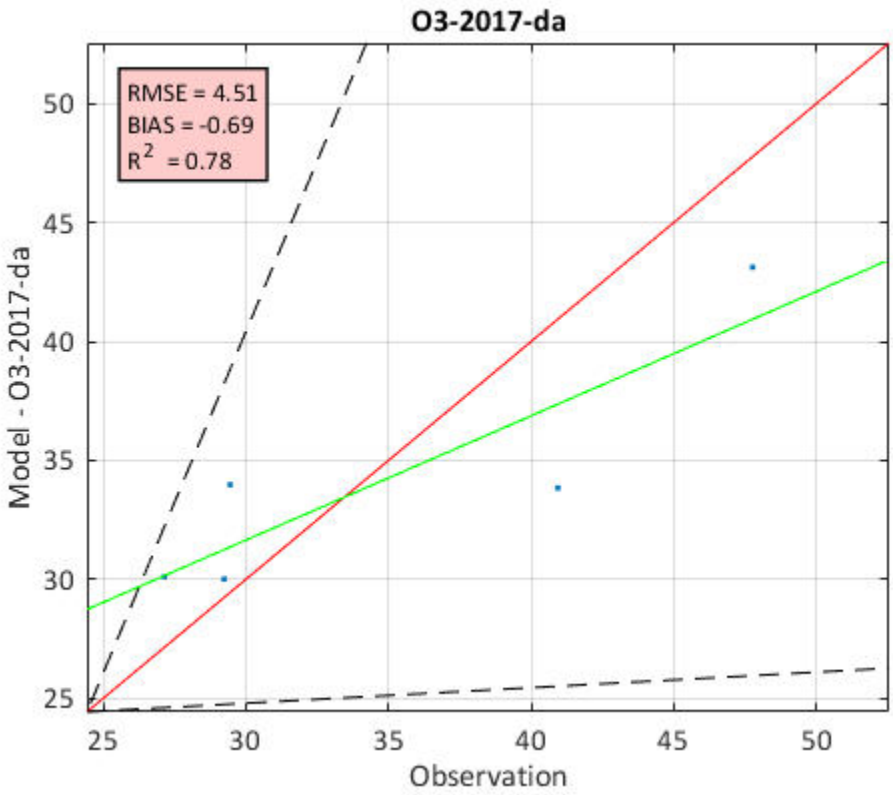


Figure 32 Comparison of modelled (using RIO) and actual ozone concentrations for 2017

Model performance and validation: temporal scale

Using the β values based on annual average pollutant concentration, daily values of pollutants at different monitoring stations are predicted. The daily modelled values are compared with daily observations at different stations. Temporal validation of the model has been carried out by comparing the daily modelled values with the actual observations. Figure 33 shows the coefficient of correlation between daily modelled and actual pollutant concentration for $PM_{2.5}$, PM_{10} and O_3. It can be observed that for $PM_{2.5}$ most of the stations showed satisfactory correlation between model and actual values except showing low values at IGI airport. In case of PM_{10} (with less number of available stations), two stations MANDIR and R1K1PU showed good correlation between daily modelled and actual pollutant concentration, but remaining two stations showed poor correlation due to over estimation of predicted modelled values.

For O_3, the model couldn't produce satisfactory correlation between daily modelled and actual pollutant concentration showing good correlation only at R1K1PU and SHADIP but less or no correlation at IGIAI, ITO111 and CIVIL1, due to over and under estimation of modelled values.

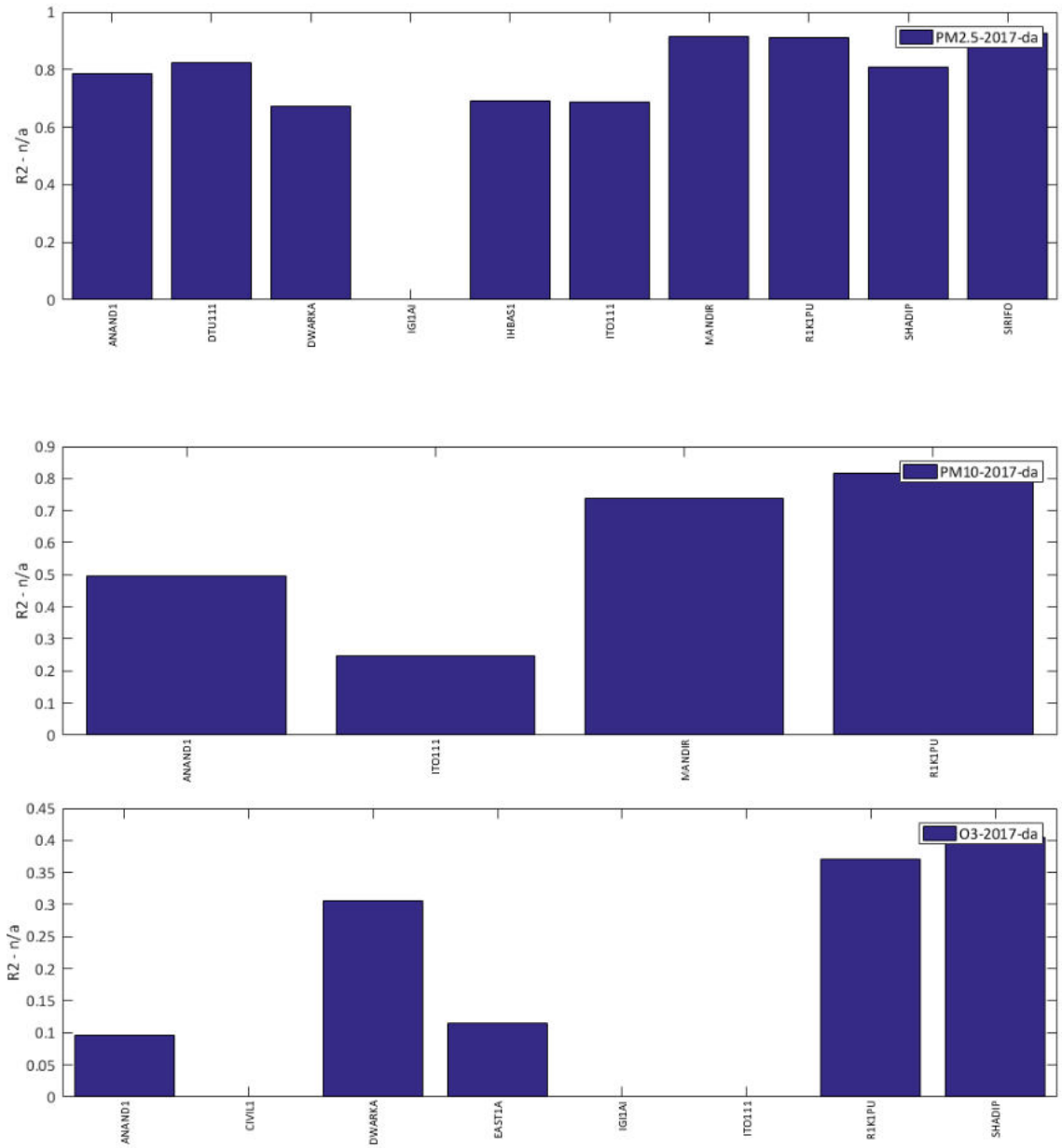


Figure 33 Coefficient of correlation between daily modelled and actual pollutant concentrations at different stations (2017)

Generation of spatial maps of air quality using validated RIO model (2017)

The validated RIO model is used to generate annual averaged maps for the year 2017 for different pollutant. The map represents the annual mean concentration PM spatially interpolated using the RIO model based on measurements at limited number of monitoring stations and the beta parameter.

PM_{10}

Figure 34 represents estimated PM_{10} spatial map for the year 2017. The small coloured circle on the map represents monitoring locations and corresponding actual monitored PM_{10} value at that station. RIO interpolation technique is applied over entire territory of Delhi suggesting that interpolation scheme is able to introduce land use based local variation in PM_{10} concentration at places where no monitoring data is available. Figure 34 shows the variation of PM_{10} concentrations ~150-500 µg/m^3 across different places in the city. This also shows that about 150 µg/m^3 is the urban background, above which local sources act and add to the PM_{10} pollutant concentrations.

The map showed highest concentration at eastern and southern part of study domain whereas lowest concentration was observed near the vicinity of "green space" area. Considering the land cover characteristic of our study domain, highest concentration was estimated near industrial and traffic bound central regions. Apart from industrial or traffic areas, higher concentration was also estimated over the south-eastern part of the city, indicating the flow of pollution in this direction due to atmospheric transport.

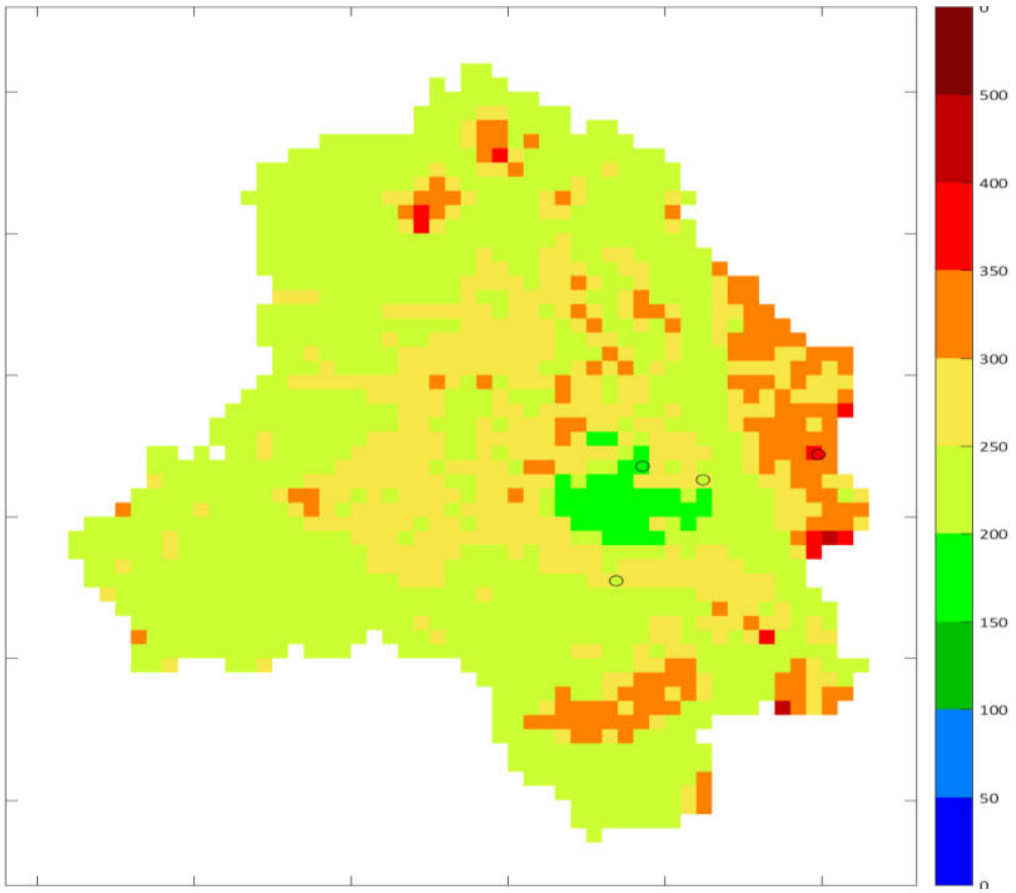


Figure 34 Spatial plot of modelled annual averaged PM_{10} concentrations using RIO for the year 2017

$PM_{2.5}$

$PM_{2.5}$ spatial maps for the year 2017 were also generated for the annual averaged. Figure 35 shows the variation of $PM_{2.5}$ concentrations ~60-250 µg/m³ across the cities. This also shows that about 60 µg/m³ is the urban background, above which local sources act and add to the pollutant concentrations.

The small coloured circle on the map represents monitoring locations and corresponding modelled $PM_{2.5}$ value. The map showed highest concentration at some of the northen parts of study domain due to presence of industrial areas. Lowest concentration was observed near the vicinity of "green spaces and water bodies. Considering the land cover characteristic of our study domain, highest concentration was estimated near the vicinity of industrial and traffic bound regions (in the central and south-eastern part of the city)

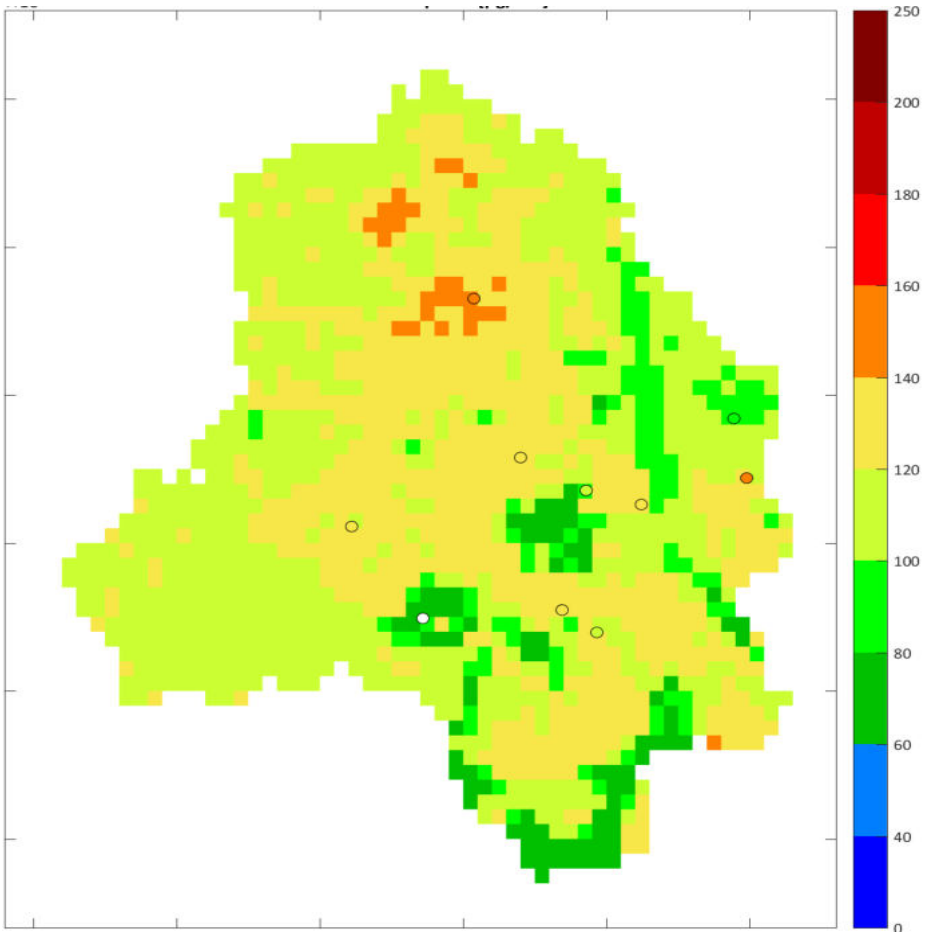


Figure 35 Spatial plot of modelled annual averaged $PM_{2.5}$ concentrations using RIO for the year 2017

Although, the RIO model performed reasonably well for preparation of spatial plots, limited number of monitoring stations remains as a constraint. In 2018, several new monitoring stations have been installed which expand the geographical coverage of monitoring data and can help in further improving the RIO model estimates. In this view, we collected data of Jan-2018 to December-2018 and again setup the model after including more number of stations in the analysis.

4.2 Spatial model results based on air quality data 2018 (Jan-Dec)

Same methodology has been applied for estimation and optimization of beta parameter.

TERI and ARAI concluded a study in 2018, were a new inventory of emissions for 2016 has been developed (Table 10). In the analysis for 2018, we have used the emissions from TERI&ARAI (2018) as starting values for ai.

Table 10 Annual Emission inventory of pollutants (kt/yr) in Delhi and NCR for 2016

SECTOR	DELHI					
	PM_{10}	**$PM_{2.5}$**	**NOx**	**SO_2**	**CO**	**NMVOC**
Transport	12.8	12.4	126.9	1.1	501.1	342.1
Industries	1.3	1.1	1.6	4.6	0.2	0.0
Power plants	5.3	3.2	11.2	23.6	3.5	0.9
Flyash ponds	0.6	0.2	0.0	0.0	0.0	0.0
Residential	2.9	2.0	3.7	0.2	61.1	12.7
Agricultural burning	0.5	0.4	0.1	0.0	2.7	0.3
Road dust	24.0	5.8	0.0	0.0	0.0	0.0
Construction	14.2	2.7				
Dg sets	0.1	0.0	0.7	0.0	0.2	0.1
Refuse burning	1.4	1.2	0.5	0.1	4.6	2.7
Crematoria	0.4	0.2	0.1	0.0	2.2	1.2
Restaurant	1.4	0.8	0.4	1.3	2.5	0.4
Airport	0.1	0.1	6.6	0.5	13.6	7.0
Waste incinerators	0.5	0.3	4.1	1.6	0.9	0.0
Landfill fires	1.8	1.5	0.6	0.1	5.8	2.2
Solvents						57.3
Coal handling at power plants	0.1	0.1				
Total	68	32	156	33	598	427

SECTOR	NCR					
	PM_{10}	$PM_{2.5}$	NOx	SO_2	CO	NMVOC
Transport	*68.6*	66.5	528.9	4.4	1750.9	886.5
Industries	288.3	127.4	85.2	556.2	620.0	27.0
Power plants	64.4	38.6	132.5	297.1	13.4	9.4
Flyash ponds	7.7	1.5	0.0	0.0	0.0	0.0
Residential	204.3	131.5	38.0	16.8	1700.3	374.1
Agricultural burning	174.1	102.2	30.6	9.0	781.1	209.2
Road dust	137.2	30.6	0.0	0.0	0.0	0.0
Construction	43.7	7.8				
Dg sets	3.7	3.2	53.0	3.5	11.4	4.3
Refuse burning	17.5	14.4	5.5	0.7	56.0	33.3
Crematoria	1.5	0.8	0.2	0.0	7.7	4.3
Restaurant	1.7	1.0	0.5	1.6	2.9	0.4
Airport	0.1	0.1	6.6	0.5	13.6	7.0

SECTOR	NCR					
	PM_{10}	$PM_{2.5}$	NOx	SO_2	CO	NMVOC
Waste incinerators	0.5	0.3	4.1	1.6	0.9	0.0
Landfill fires	1.9	1.6	0.6	0.1	6.1	2.3
Solvents						112.8
Coal handling at power plants	1.6	1.0				
Total	1017	528	886	892	4964	1671

Source TERI and ARAI (2018)

The broad landuse classes used in this analysis remain same as in Table 11.

Table 11 Broad landuse categories used in the model

S.No.	Landuse class
1	a1 = Residential, Drain
2	a2 = Industrial, Commercial
3	a3 = Green Area
4	a4 = Cultivated
5	a5 = Open Space, Others
6	a6 = River, Water body, Canal, Lake, Island

The area under each of the landuse in the city is estimated and presented in Table 12.

Table 12 Area under each of the landuses in Delhi

Land cover	Emission class	Area (km^2)	Percentage
Canal	a6	0.870302	0
Commercial	a2	13.4998	1
Cultivated Land	a4	551.17	37
Drain	a1	6.619793	0
Green Area	a3	84.9826	6
Industrial Area	a2	28.26872	2
Island	a6	1.173269	0
Lake	a6	0.2615	0
Open Space	a5	58.02503	4
Others	a5	40.61282	3
Residential Area	a1	680.866	46
River	a6	13.1588	1
Water Body	a6	1.887822	0

The area under different clubbed landuse categories as per Table 12 is shown in Table 13. Evidently 46% of area is under residential, and 37% is cultivable land. Only 3% area is under industrial/commercial landuse.

Table 13 Area under clubbed landuse category in Delhi

Emission class	Area (km^2)	Percentage
a1	687.48582	46
a2	41.768515	3
a3	84.982596	6
a4	551.170024	37
a5	98.637843	7
a6	17.35169	1

Establishing land use β parameter

Spatial driver is constructed from land cover information. For initial estimates of β, initial pollutant coefficient a_i is derived from source emission estimates for Delhi (Table 14). Emission estimates are used as the initial estimates for ai, as they provide the initial guess for the local characteristics of air quality. The weights for the different land cover classes have been initially estimated from the emission inventory for Delhi, by scaling the land cover weights of different landuses with the value for residential area set at unity. The estimate of initial ai values is shown below in Table 14.

Table 14 Initial ai values derived using sectoral emission estimates of Delhi city

a		Sectors for initial weights	Emission total $PM_{2.5}$	Scaled	Scaled with area
1	Residential area	S1+S2+S6+S8	24.1	1	1
2	Industrial + Commercial	S3+S5+S9	7.5	0.31	5.12
3	Urban Green	-	-		0
4	Cultivated land	S7+S4	0.4	0.0016	0.021
5	Open space	-			0
6	Water bodies	-			0
		SUM:	32		

Where,

Sector	Label
S1	Transport
S2	Domestic
S3	Diesel Gen
S4	Brik Kilns
S5	Industries
S6	Construction
S7	Waste Burning
S8	Road Dust
S9	Power Plant

Optimization of β parameter

For the year 2018 (Jan-Dec), various optimization scenarios have been run to setup the

β parameter for various monitoring stations. These scenarios are explained below and the land use optimization values are shown in Table 15.

Table 15 Description of scenarios used for model optimisation purposes

LC1	LC2	LC3	LC4	LC5
Residential: 1	Residential: 1	Residential: 1	**Same** land use optimization as LC3 with manual adjustments for Anand Vihar and Pusa stations	**Same** land use optimization as LC3 with exclusion of Anand Vihar and Pusa stations
Industry: 4.6	Industry: 4	Industry: 1.9		
Urban Green: 0.1	Urban Green: 0.05	Urban Green: 0.05		
Cultivated land: 1.25	Cultivated land: 0.5	Cultivated land: 0.5		
Open Space: 0	Open Space: 0.15	Open Space: 0.15		
Water: 0	Water: 0	Water: 0		

LC1: Spatial driver for PM

First optimization has been carried out, which led to a higher contribution for cultivated land and non-zero contribution for urban green. Two 'odd' stations Okhla and Najafgarh are quite affected by industrial land use contributions, but is not reflected by higher PM concentrations. Possibly these industrial locations are not linked to significant local PM emissions, or these industrial emissions are stack emissions and do not have large impact at ground level. The stations are left out to optimize the land-use weights and are added with lower beta in the stations file (beta 0.6).

Optimization pushes the contribution for industry and cultivated land higher. Given the impact of agricultural fires and possible impact of linked land uses such as brick kilns, the contribution for cultivated land has been allowed to be slightly higher as the residential sector. The trend of PM against beta (based on land cover) is determined with this set of stations, including the 2 adjusted stations.

LC2: Spatial driver for PM

This was a further variation on LC1, with lower weight for cultivated land, and slight adjustments for other land cover classes based on an optimization step. Two 'odd' stations Okhla and Najafgarh are left out, as done for LC1. In this scenario, there is slightly lower weight for industry, lower weight for cultivated land and a minor contribution for open space and urban green. The trend of PM against beta (based on land cover) is determined with this set of stations, including the 2 adjusted stations.

LC3: Spatial driver for PM

This was a further variation on LC2, where a third optimization is performed with still lower contributions from industry, still neglecting stations Okhla_ and Najafgarh. The trend of PM against beta (based on land cover) determined with this set of stations, including the 2 adjusted stations.

LC4 Spatial driver for PM

This scenario uses identical land use optimization as in LC3. In addition to this, the two outlier stations, Anand Vihar and Pusa have been manually adjusted to beta-values better in

line with the observed trend. Adjusting these stations has an impact of the fit of PM concentrations against the spatial driver beta. The trend of PM against beta (based on land cover) is determined with this set of stations, including the 4 adjusted stations.

LC5 Spatial driver for PM

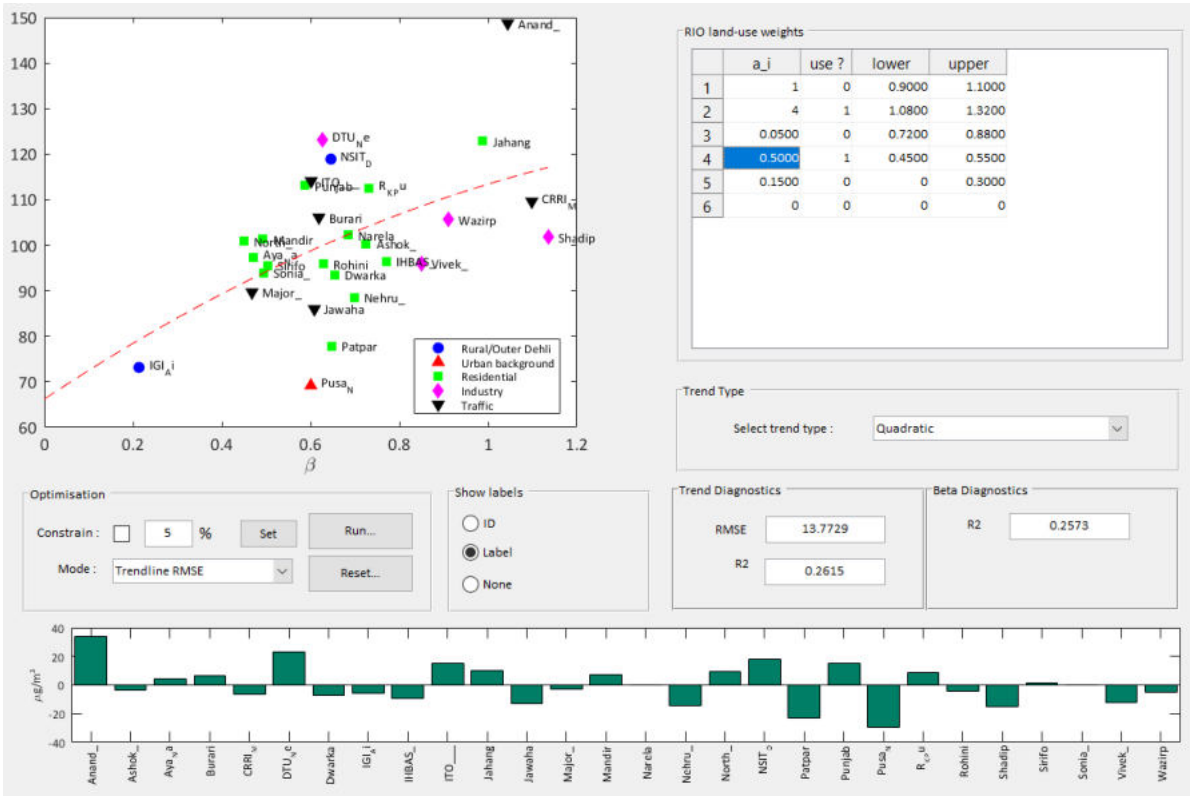
Identical land use optimization as LC3 and 4, but completely neglecting stations Anand Vihar and Pusa New Dehli. The trend of PM against beta (based on land cover) is determined with this set of stations, excluding Anand and Pusa.

Figure 36 shows the plot of annual average concentration of $PM_{2.5}$ of different stations with β parameter in the LC1-LC4 scenario.

LC1


RIO land-use weights
a_i use ? lower upper
1 0 0.9000 1.1000
4.6000 1 1.0800 1.3200
0.1000 0 0.7200 0.8800
1.2500 1 0.4500 0.5500
0 0 0 0
0 0 0 0
Trend Type
Select trend type : Quadratic
Optimisation
Constrain : 5 % Set Run...
Mode : Trendline RMSE Reset...
Show labels
ID
Label
None
Trend Diagnostics
RMSE 13.6158
R2 0.2783
Beta Diagnostics
R2 0.2759
Rural/Outer Dehli
Urban background
Residential
Industry
Traffic
β
μg/m³

LC2


RIO land-use weights
a_i use ? lower upper
1 0 0.9000 1.1000
4 1 1.0800 1.3200
0.0500 0 0.7200 0.8800
0.5000 1 0.4500 0.5500
0.1500 0 0 0.3000
0 0 0 0
Trend Type
Select trend type : Quadratic
Optimisation
Constrain : 5 % Set Run...
Mode : Trendline RMSE Reset...
Show labels
ID
Label
None
Trend Diagnostics
RMSE 13.7729
R2 0.2615
Beta Diagnostics
R2 0.2573
Rural/Outer Dehli
Urban background
Residential
Industry
Traffic
β
μg/m³

LC3

LC4

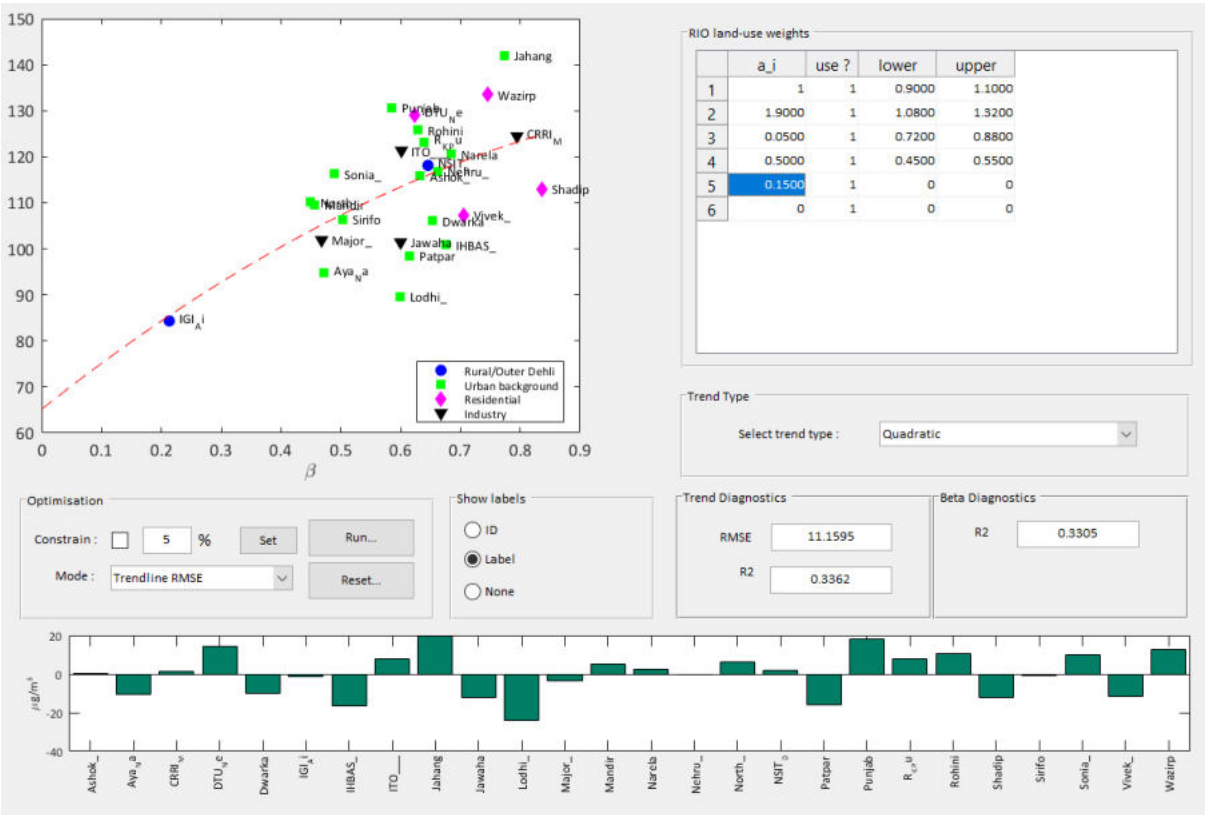


Figure 36 Optimization of the weights of the land cover classes to improve the trend of the annual average $PM_{2.5}$ concentrations for the stations in Delhi against their beta-values (LC1-LC4).

Calculation of a_i parameters applied for RIO

The intelligent spatial interpolation model RIO has been configured based on land use information. To apply land use as spatial driver the data are converted into a single parameter beta. As first step, land use classes are lumped into 6 lumped classes and matched with the available emission inventory. The sectoral emissions are used as first approximation to derive the weight of each lumped land use class, taking into account area of Delhi belonging to these land use classes.

The calculation of the beta parameter is calculated for the surroundings of each station and for each RIO grid cell using the formula:

$$\beta = \log\left[1 + \frac{\sum_i a_i \times n_{\mathrm{RCL}i}}{\sum_i n_{\mathrm{RCL}i}}\right].$$

With a_i the weight per lumped land use class and n_{RCLi} the area covered by that lumped class in that RIO grid cell.

As a next step, the weights a_i are optimized to have clear trend between the beta-values per station and the long term average concentrations at these stations. The configuration of the RIO model includes a routine to optimize a_i parameters for minimal RMSE. However, only after validation on the final concentration outcomes of the RIO model, decisions on the most appropriate a_i weights are made.
Several different approaches to optimize these weights a_i have been tested and validated against the station observations using a leaving-one-out approach to decide on the best set of a_i values.

As explained above, several sets of lumped land use class weights have been tested and validated. The final weights have been chosen based on the outcomes of these validations and the export knowledge on the local air quality situation in Delhi.

The optimization model runs to derive ai values for $PM_{2.5}$ for LC1 to LC4 scenarios. Different optimization scenarios led to generation of different beta-values for monitoring stations and other grids in the study domain. The differences in beta-values can be observed in Figure 37 for the LC1 and LC2 scenarios. The LC2 scenarios clearly show lower beta-values for outer Delhi region suggesting lower local emissions in these regions.

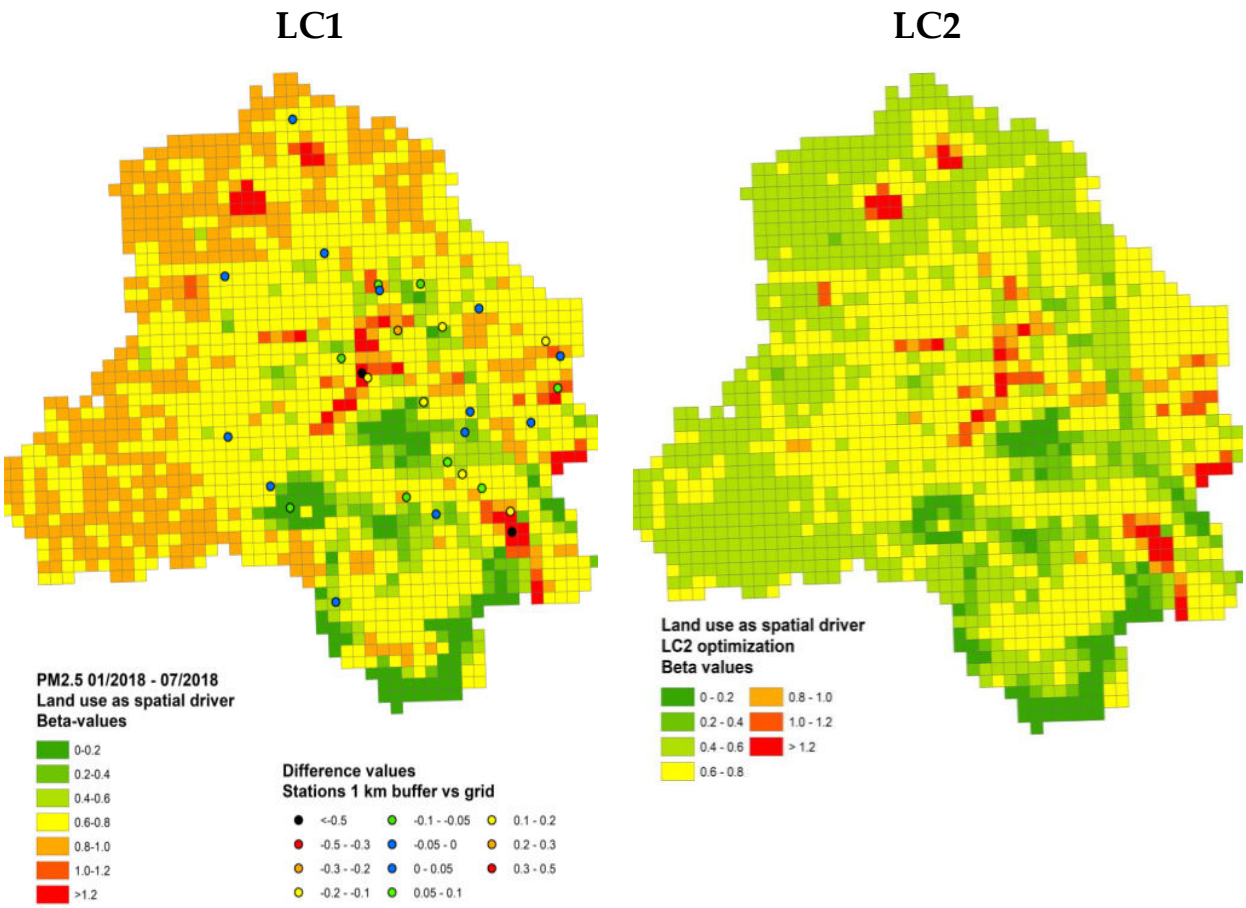


Figure 37 Beta parameter values for the LC1 and LC2 scenarios

The scenario analysis reveals that LC4 is the best optimisation scenario, which is able to predict the $PM_{2.5}$ concentrations most effectively. It can be seen from Figure 36 that LC4 scenario shows highest correlation compared to other scenarios with a RMSE of 11.15. Therefore we have taken LC4 scenario for validation purpose. It was observed that for high β values the concentration was also high at residential and traffic locations, showing the local activities leading to local impacts on air quality.

4.2.1 Validation of spatial model for 2018 (Jan-Dec)

Validation of any model is an essential step before using the model for any further processes. Spatial and temporal validations were carried out for the modelled results under different scenarios.

Spatial validation

Spatial validation of the RIO model results has been carried out for $PM_{2.5}$ for the entire year 2018 by considering leaving one out approach. The exercise was carried out using the best performing LC4 scenario. Such approach was carried out for all the stations in which model results (after leaving out 1 particular station for estimation of β trends) were compared with the observed concentration value. Figure 38 shows the correlation plot between modelled and observed concentrations for different monitoring stations in the city. It is observed that modelled values for most stations were close to their observation however, some stations

showed over estimation of predicted modelled values. The coefficient of correlation (R^2) was found to be 0.45. Overall, the RMSE value was 11.78 while bias remaining at -0.54.

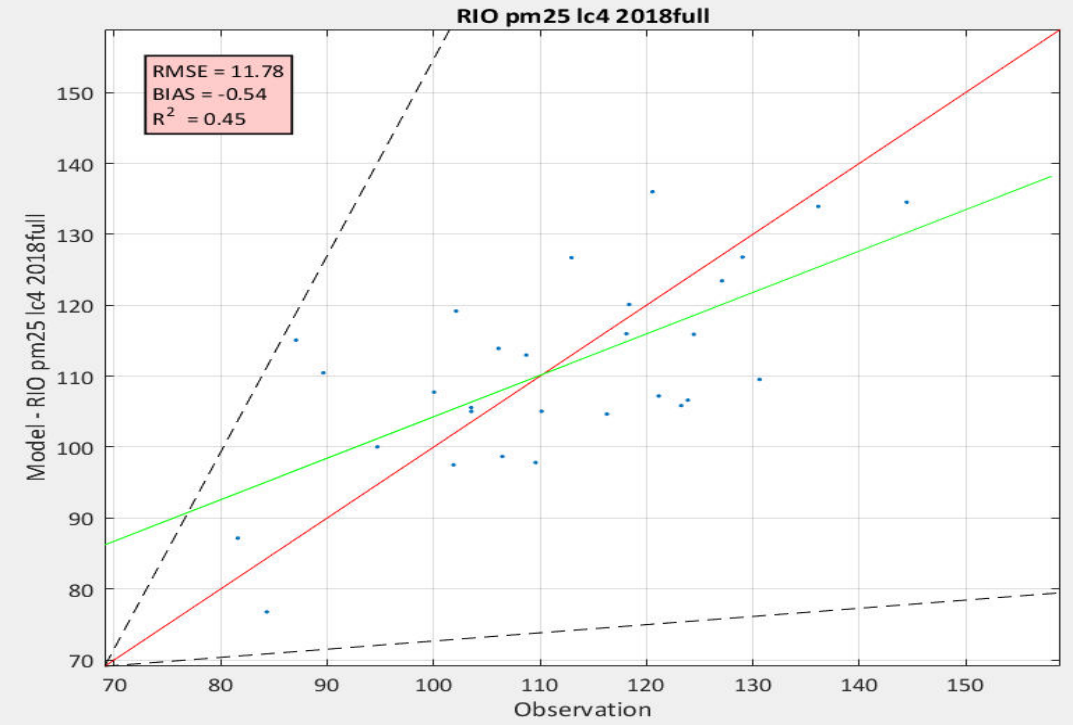


Figure 38 Comparison of modelled and observed $PM_{2.5}$ concentrations at different stations in Delhi under LC4 scenario (2018)

Temporal validation

The model developed using β values and annual average pollutant concentration is used to predict daily pollutant concentration under LC4 scenario. The daily modelled $PM_{2.5}$ values are compared with observations at different stations. For comparative evaluation, three performance metrics were chosen-coefficient of correlation, bias and RMSE.

Figure 39 shows the three plots under LC4 scenario showing the coefficient of correlation, bias and RMSE values between daily modelled and actual pollutant concentration for different stations. It was observed that most stations out of 29 locations show good correlation with daily observed concentration having R^2>0.6. Only 1 station (Aya Nagar) has shown correlation value of less than 0.5. This confirms satisfactory performance of the model for predicting daily variations of concentrations at most of the stations. Most stations also showed less bias for estimated modelled values when compared to daily observed value. The bias remained less than ±20 μg/m^3 for most stations. In the LC4 scenario, only 2 stations-LODHI and PUNJAB showed bias more than ±20 μg/m^3. The results from RSME also showed low values for most of the locations. Only 7 out of 29 stations were observed to have higher RMSE values ranging between of 40-60 μg/m^3. Only two stations, Aya Nagar and IHBAS showed RMSE values higher than 50 μg/m^3. Overall, this can be concluded that for most of the stations the RIO spatial model is performing satisfactorily for entire year 2018 (Jan-Dec) and hence can be used for prediction of values at other locations in the city and generation of satisfactory spatial air quality map.

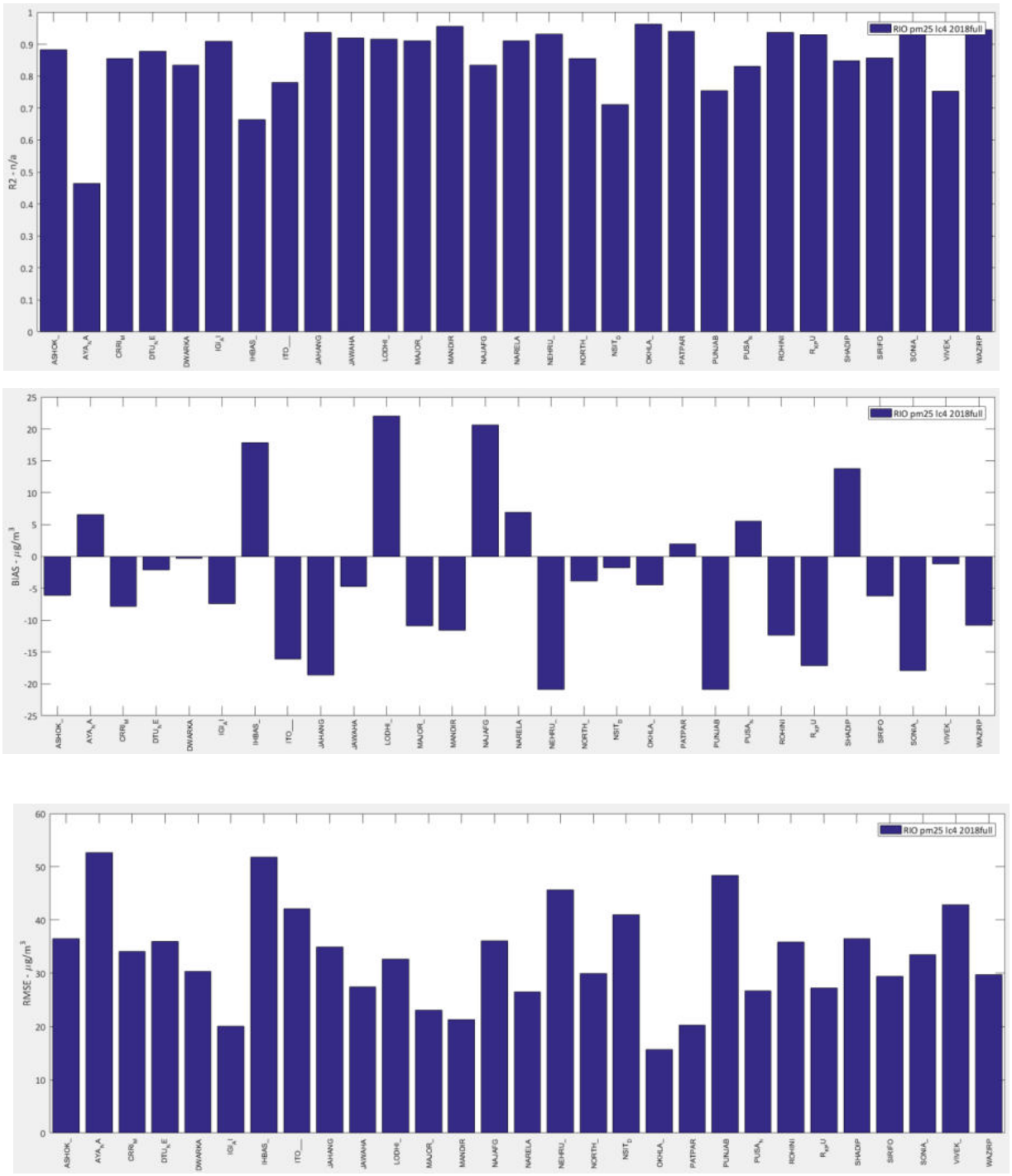


Figure 39 Plots under LC4 scenario showing the coefficient of correlation, bias and RMSE values between daily modelled and actual pollutant concentration for different stations in 2018

Figure 40 shows the plots of ratios of a) bias to standard deviation, and b) RMSE to standard deviation. It is evident that the model has very low bias and high precision. Most values show low RMSE/σ values also for most of the stations. Only 2 stations out of 29 are beyond the satisfactory ranges (as depicted by green circles) of these two ratios.

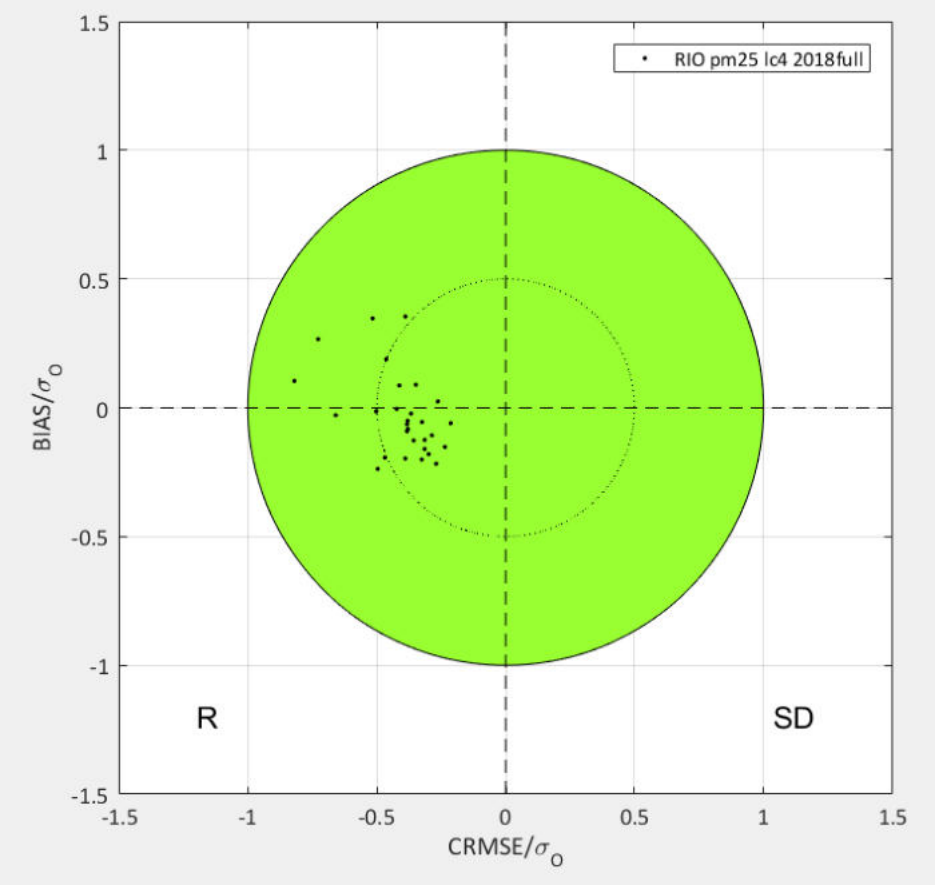


Figure 40 Ratios of a) bias to std. deviation and b) RMSE to std. deviation between the modelled and predicted values of PM2.5 at different stations in 2018.

Each dot represents one station, the value on x-axis is the ratio of the centered root mean square error CRMSE over the standard deviation of the observations and the y-axis the ratio of the bias over the standard deviation of the observations (introduction of the target plot for model performance evaluation in Pederzoli et al. 2012, IJEP 50, 175-189, Performance criteria for the benchmarking of air quality model regulatory applications: The 'target' approach).

4.2.2 Generation of spatial plots $PM_{2.5}$

Once the model is validated, this can be reliably used for prediction of pollutant concentrations at the locations, where there is no monitoring stations available. The predicted values can then be used to generate spatial air quality maps for various pollutants on daily, monthly or annual basis.

Figure 41 presents the spatial $PM_{2.5}$ map for year 2018 based on the most suitable LC4 scenario, for the study domain i.e. Delhi. The small coloured circles on the map represent monitoring locations. This can be observed that the colour in the circle generally corresponds to the $PM_{2.5}$ values estimated by the RIO model value. The maps also suggest that RIO interpolation technique is also able to introduce land use based local variation in estimated $PM_{2.5}$ concentration at places where no monitoring data is available.

LC4 scenario is a scenario with readjustment for the two locations namely, Anand Vihar and Pusa road. The explanation for this readjustment specifically for these 2 locations is as follows. RIO is an intelligent interpolation model which uses spatial information to adjust (detrend) the observations of pollutant concentrations prior to interpolation. After the interpolation, the trend is reapplied to create the resulting concentration maps. During this procedure, it is assumed that the observations monitored at a specific station are representative for its surroundings. Some stations can however be heavily influenced by local sources and are not representative for the concentrations in the surroundings. An example is a kerbside station like Anand Vihar, which is significantly influenced by local traffic emissions. An essential step during the RIO configuration is the determination of the relation between the spatial driver and average concentrations observed at the stations. If a station is not representative for its surroundings, the model results can be improved by either adjusting the spatial driver for a specific station or neglecting the station. During the RIO configuration for Delhi, stations which are characterized by industrial land use in their surroundings but do not show increased PM concentration levels have been manually adjusted to rather reflect urban background stations. For a hotspot location like Anand Vihar, we can consider this an outlier and leave the station out or adjust its spatial driver information as well. For an outlier with low concentrations, it can be considered that the level of detail of the spatial information is insufficient to explain the difference between such a location and other urban background locations in Delhi. As all stations in Delhi are mainly characterized by a domination of residential land use, a station like Pusa New Delhi is more representative for locations with less urban influence. Therefore, the land use parameter for this station could be manually adjusted to lower values. Ideally, more detailed information is available in the future which can explain for the difference within urban background stations and avoiding the need of manual adjustments.

The maps also suggest that RIO interpolation technique is also able to introduce land use based local variation in estimated $PM_{2.5}$ concentration at places where no monitoring data is available. It could be observed from Figure 41 that the highest concentration (120-135 $\mu g/m^3$) was estimated at some north-western and south-eastern, central part of study domain indicating hot spots mostly characterized by urban and industrial land cover. High contribution with concentration value of 105-120 $\mu g/m^3$ was also estimated at some northern, eastern and south-eastern region of the map mostly covered by urban/residential area or traffic location. Model still estimated high contribution (90-105 $\mu g/m^3$) at western part of study domain indicating the impact of background concentrations i.e. the influence of emissions generated outside (upwind) regions in Delhi's $PM_{2.5}$ concentrations. As expected, the lowest concentrations (75-90 $\mu g/m^3$) were estimated at green cover and open space area. This scenario provided the closest estimates to the actually observed $PM_{2.5}$ concentrations at most monitoring stations in Delhi.

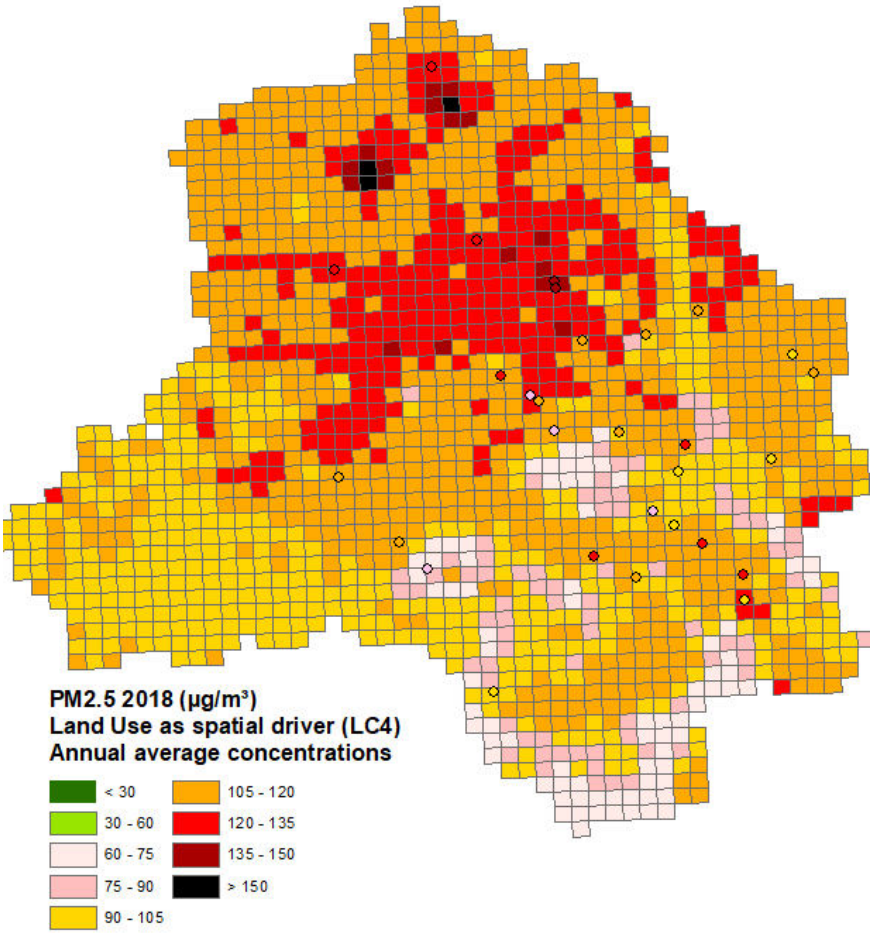


Figure 41 Annual average PM2.5 map for 2018 using RIO with land cover information as spatial driver (LC4) on daily average basis.

4.2.3 Optimization analysis of $PM_{2.5}$ with Population Density

An attempt has also been made to assess the model optimization with population density as a parameter instead of landuse categorization. Figure 42 represents population density map of our study domain utilizing population density data from GHS 2015. In the map, densely populated regions were observed at central, eastern and some south-eastern part of the study domain due to impact of urban/residential land cover pattern. The region with less population density was estimated at green space (1500-1750) and rural/cultivated area (<100).

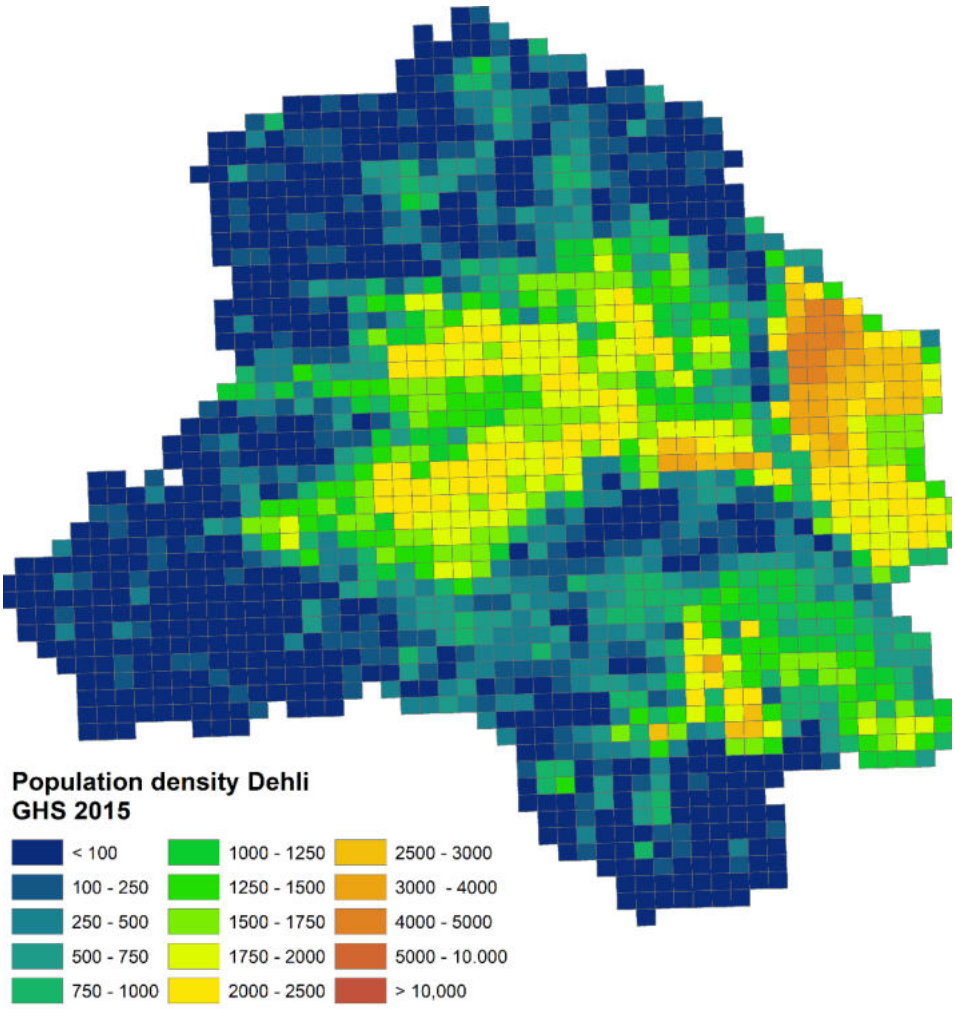


Figure 42 Population density map of Delhi (2015)
Data source: GHS 2015

The figure 43 represents correlation between $PM_{2.5}$ concentrations and log(1+pop) for all stations within our study domain. Average population density was estimated within a 1km buffer around the stations. No significant correlation has been observed between the two parameters, and hence population density has not been considered as the parameter for deriving beta-values.

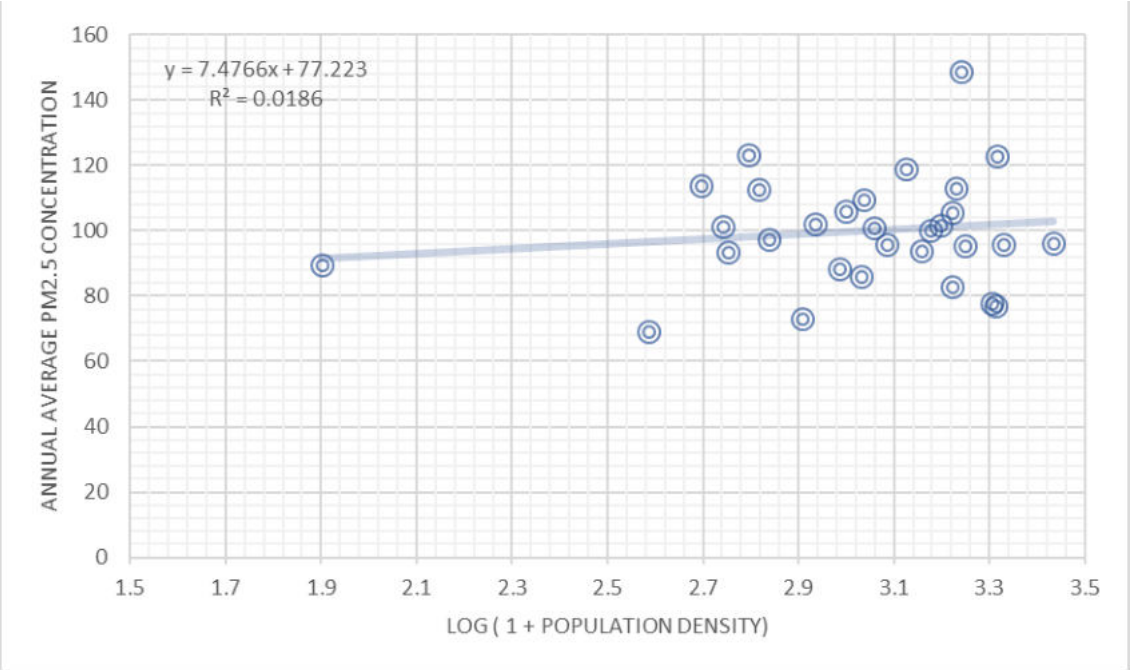


Figure 43 Correlation of $PM_{2.5}$ concentrations with population density in the buffer around the monitoring stations.

Although Population density didn't turn out to show good correlations with $PM_{2.5}$, model optimizations were performed and spatial maps was prepared using this as the spatial parameter to generate beta-values. The figure 44 represents estimated $PM_{2.5}$ map for year 2018 based on population, as spatial driver for construction of $PM_{2.5}$ map on the study domain. The small coloured circle on the map represents monitoring locations and corresponding estimated modelled $PM_{2.5}$ value. RIO interpolation technique is also applied over the entire region suggesting that interpolation scheme is able to introduce land use based local variation in estimated $PM_{2.5}$ concentration at places where no monitoring data is available. In the map, highest concentration (100-120 $\mu g/m^3$) was estimated at northern, central and eastern part of study domain indicating highly populated regions due to residential land cover pattern. The remaining part of study was estimated with less $PM_{2.5}$ concentration due to open space and cultivated land cover characteristics. It is evident that population density it not able to predict well the concentrations at the monitoring locations in Delhi.

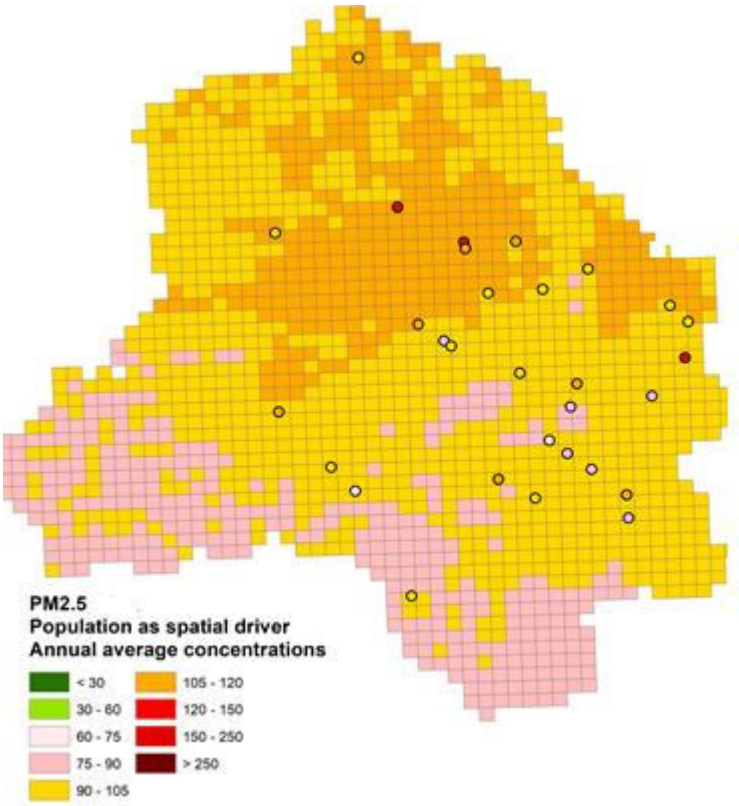


Figure 44 Spatial map based on population density as the parameter

The model values were characterized by high RMSE and BIAS values, and also showed very low spatial correlations (Figure 45).

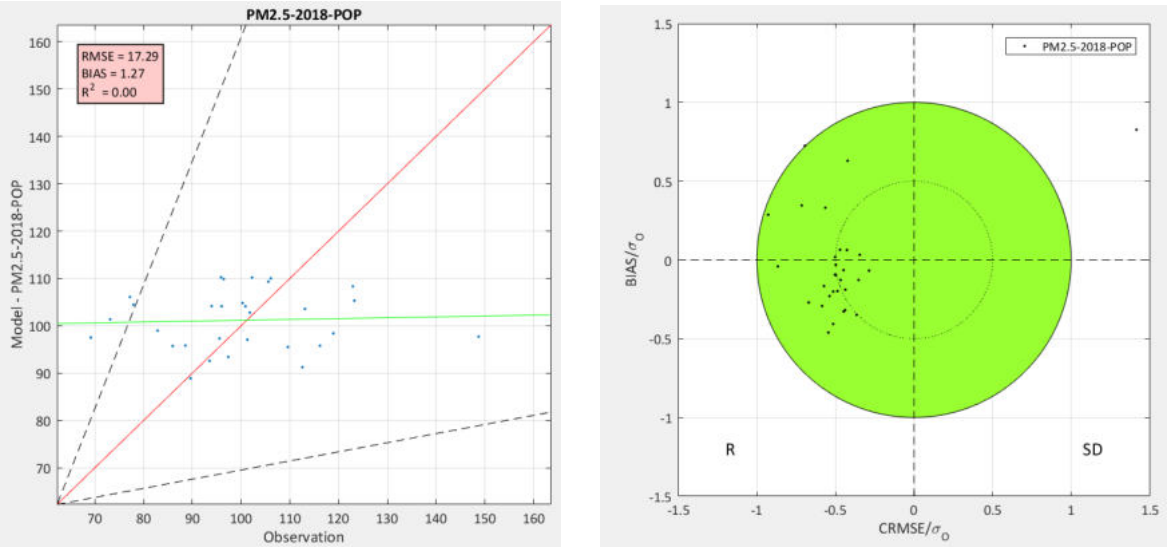


Figure 45 Performance of model using population density as the spatial driver – a) comparison of modelled and actual $PM_{2.5}$ concentrations, b) Plots of ratios of Bias and RMSE to standard deviations

4.2.4 Comparison of ordinary krigging interpolation with RIO model results

Spatial maps can also be prepared using simple krigging method of interpolation between the pollutant concentrations values between different stations in Delhi. Figure 46 shows the annual $PM_{2.5}$ concentrations for the period Jan-Dec 2018 predicted using a) RIO model for LC3 scenario b) normal Krigging interpolation method. A comparison of the two maps evidently depicts the advantages of the RIO model as it is able to highlight the local variations in air quality based on local characteristics. Standard techniques have to rely only on the monitoring data themselves and their relative distances and correlations. RIO model can deal with the urban stations in a much better way by taking into account the local characteristics and assign lower and higher values to the relevant areas.

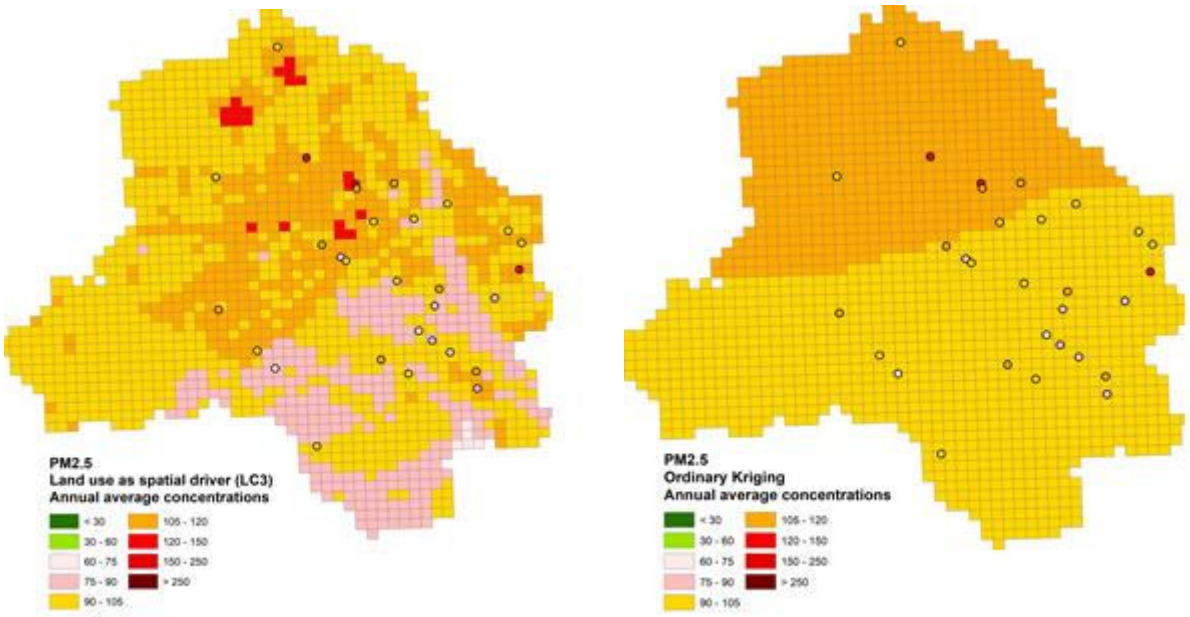


Figure 46 Comparison between results of RIO model and ordinary krigging method

4.2.5 Spatial model results for PM_{10}

Similar to optimization analysis of $PM_{2.5}$, different landuse are used as spatial driver in our study domain to derive beta values for all 27 stations. Estimation of daily average concentration of PM_{10} for year 2018 (Jan-Dec) as a function of β parameter derived using LC2 optimization scenarios was in terms of fitting of the curve. LC2 scenario Beta values computed using all the three scenarios was used for prediction of values at the locations where no monitoring stations is available. Figure 47 shows the model optimization runs for the LC2 scenario for PM_{10}. The trend yielded fit that shows r^2 value of 0.39, and RMSE of 27.15. The model estimated highest concentration (>300 µg/m³) at station Wazirpur and Jahangir Puri characterized by industrial and residential land cover, respectively. Overall, the plot yielded observation that for high β values the concentration was also high at residential, industrial and traffic locations showing the local activities leading to local impacts on air quality.

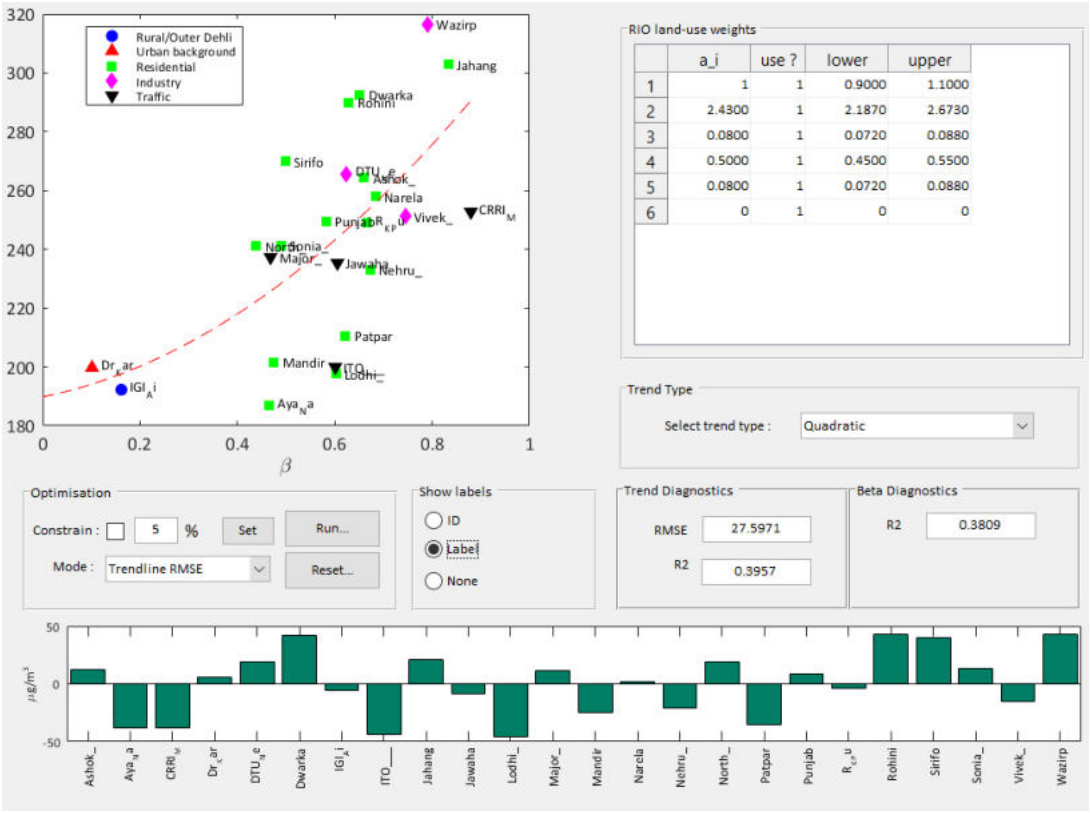


Figure 47 Model optimization run for the LC2 scenario for PM10 (2018)

The PM_{10} trend show a clear increase in PM levels with increasing beta-parameter, the remaining scatter of the stations around the trend line cannot be further optimized based on the land cover information as many monitoring stations in Delhi have a similar (urban) land cover foot print but still show fairly different annual average PM_{10} concentrations.
Four monitoring stations have been neglected for the optimization of the trend for PM_{10} as a function of beta values, namely Anand Vihar, Okhla_, Najafgarh, and Pusa New Delhi. These stations have been added with manually adjusted beta values as their observations are not representative for the land cover in their vicinity.

Model validation was also carried out for PM_{10} concentrations before generation of spatial maps.

Model validation: PM_{10}

The optimized model for PM_{10} needs to be validated on both spatial and temporal scales.

Spatial Validation

Spatial validation of the model results for PM_{10} during year 2018 was carried out by considering leaving one out approach using LC2 as spatial driver. Such approach was carried out for all the stations in which model results (after leaving out 1 particular station for estimation of β trends) were compared with the observed concentration value. Figure 48 shows the results of spatial validation under LC2 scenario. It was observed from the correlation plot that optimized modelled values for most stations were close to their observation however, few stations showed over estimation of predicted modelled values. LC2 scenario is able to introduce best correlation fit between PM_{10} and beta values based on landuse characteristics of the different monitoring stations. The coefficient of correlation was found to be 0.60 whereas in case of 2018(Jan-Jul) model run the R^2 value was 0.61, producing not much difference in model output. Overall, the RMSE value was 27.25 while bias remaining at -1.56.

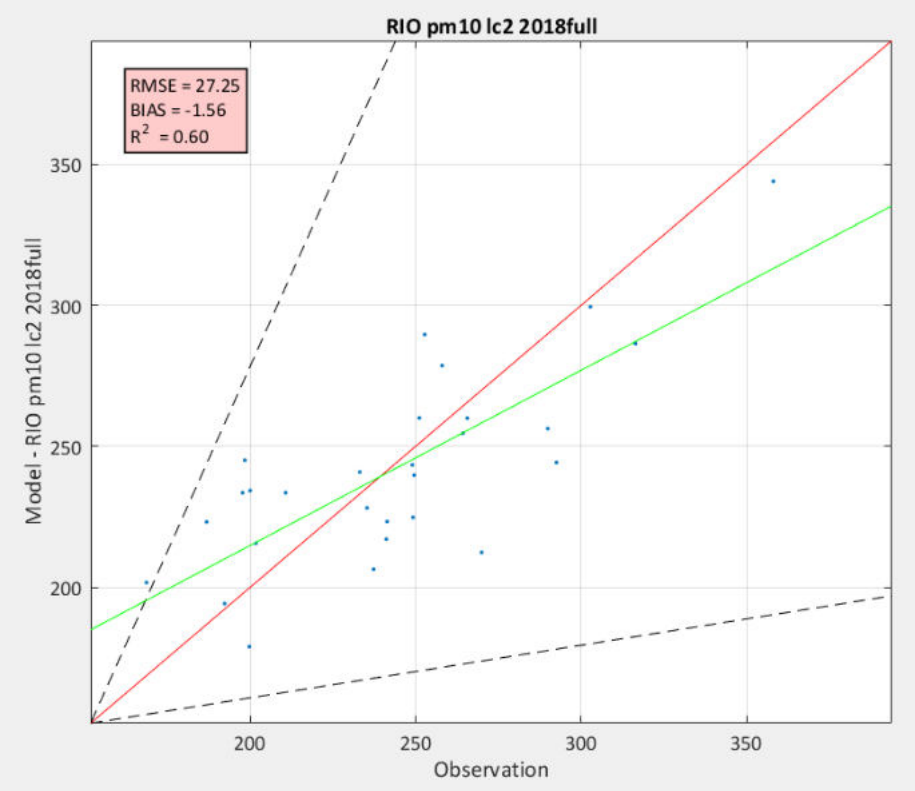


Figure 48 Model validation for PM10 under LC2 scenario - Spatial

The validation proves the potential of the RIO model to capture the PM_{10} gradients in the city of Delhi well. The remaining scatter can only be further addressed by using more

detailed spatial information or adding a dispersion model to better capture local effects at an increased resolution.

Temporal validation

Along with spatial validations, temporal validations were also carried out. The model developed using β values and annual average pollutant concentration is used to predict daily pollutant concentration for PM_{10} under LC2 scenario. The daily modelled PM_{10} values are compared with observations at different stations. For comparative evaluation, three performance metrics were chosen-coefficient of correlation (R^2), bias and RMSE.

Figure 48 shows the three plots under LC2 scenario showing the coefficient of correlation, bias and RMSE values between daily modelled and actual pollutant concentration for different stations. It was observed from Figure 49 that all the stations showed strong temporal correlation between the stations with daily observed concentration having R^2>0.8. The estimated output shows satisfactory performance of the model for predicting daily variations of concentrations at different stations. Most stations also showed less bias for estimated modelled values, when compared to daily observed value. The bias remained less than ±20 μg/m^3 for most stations, however, some of stations showed bias more than 20μg/m^3. The results also showed low RMSE values for most of the locations, however, except for two stations where higher RMSE values >80 μg/m^3 were observed. Temporal RMSE and BIAS values range between 30 and 90 μg/m^3 and -60 to 40 μg/m^3, respectively. Overall, this can be concluded that for most of the stations the RIO spatial model is performing satisfactorily for entire year 2018 (Jan-Dec) for PM10 and, hence, can be used for prediction of values at other locations in the city and generation of spatial air quality maps.

.

Figure 49 Temproal validation of daily modelled concetraitons with actual values at different stations under LC2 scenario

Figure 50 shows the plots of ratios of a) bias to standard deviation, and b) RMSE to standard deviation. It is evident that the model has very low bias and high precision. Most values show low RMSE/σ values for most stations.

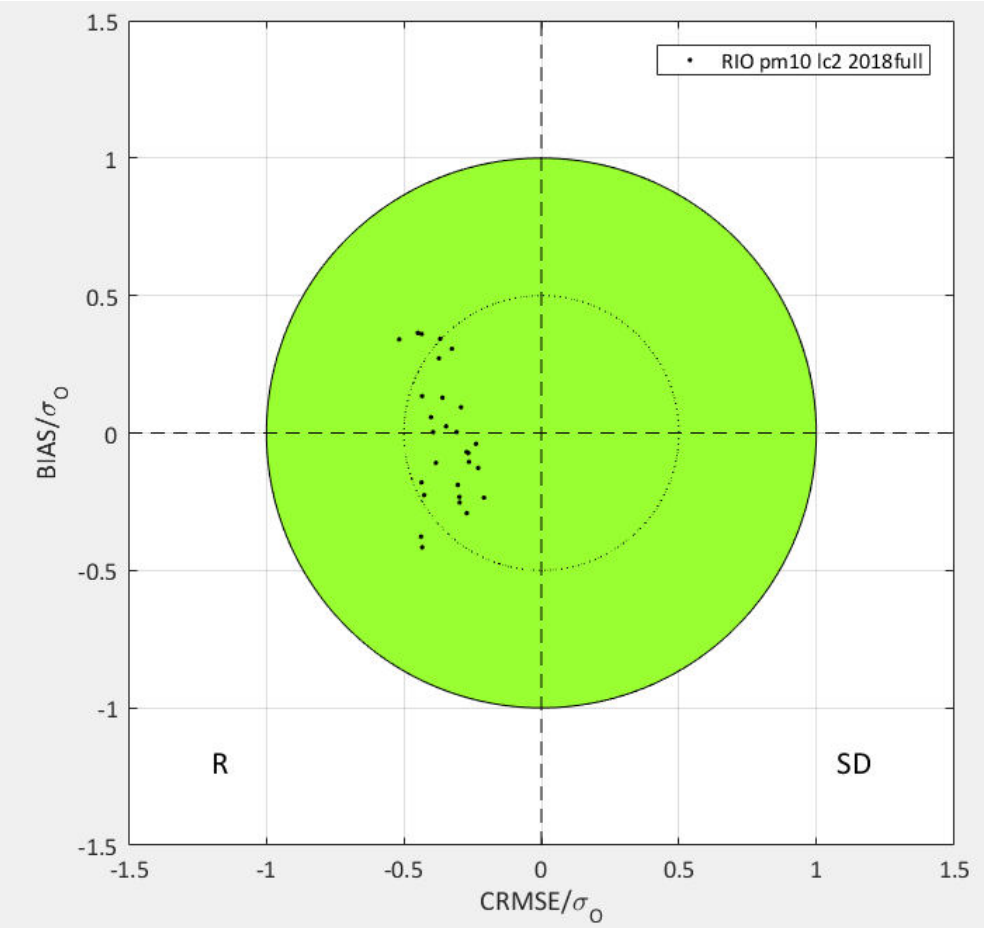


Figure 50 Ratios of a) bias to std. deviation and b) RMSE to std. deviation between the modelled and predicted values of PM10 at different stations in 2018**.**

Each dot represents one station, the value on x-axis is the ratio of the centered root mean square error CRMSE over the standard deviation of the observations and the y-axis the ratio of the bias over the standard deviation of the observations (introduction of the target plot for model performance evaluation in Pederzoli et al. 2012, IJEP 50, 175-189, Performance criteria for the benchmarking of air quality model regulatory applications: The 'target' approach).

4.2.6 Generation of spatial PM_{10} maps for Delhi

Once the model is validated this can be reliably used for prediction of pollutant concentrations at the locations, where there is no monitoring stations available. The predicted values can then be used to generate spatial air quality maps for various pollutants on daily, monthly or annual basis. Figure 51 presents the spatial PM_{10} map for year 2018 (Jan-Dec) based on LC2 scenario, for the study domain. The small coloured circles on the map represent monitoring locations. This can be observed that the colour in the circle generally corresponds to the PM_{10} values estimated by the RIO model value. The map also suggests that RIO interpolation technique is able to introduce land use based local variation in estimated PM_{10} concentration at places where no monitoring data is available. It could be observed from the map that highest concentration (300-350 $\mu g/m^3$) was estimated at some

northern, central and south-eastern part of study domain indicating hot spots mostly characterized by urban and industrial land cover. High contribution with concentration value of 275-300 μg/m^3 was also estimated at few northern and central region of the map mostly covered by urban/residential area or traffic location. Model still estimated high contribution (250-275 μg/m^3) at central part of study domain characterized by urban/residential area or traffic bound location. The remaining part of study region i.e northern and southern area was estimated with high concentration (225-250 μg/m^3). However, green cover and open space area was still estimated with high concentration (175-200 μg/m^3). Even considering low β value during optimization, the remaining part of study domain with cultivated land cover region was also estimated with high concentration suggesting the impact of agriculture crop burning or background contributions from outside of Delhi.

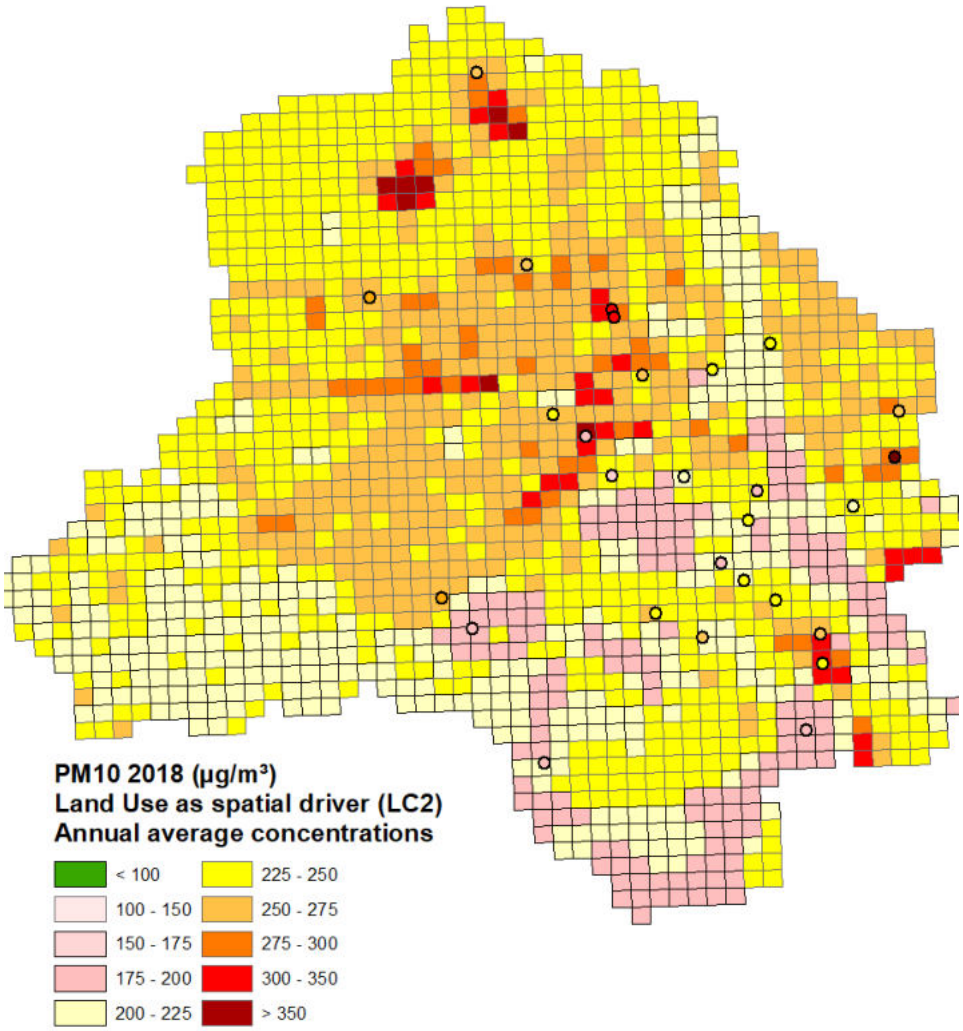


Figure 51 Spatial maps for PM_{10} for 2018 under LC2 scenario

4.3 Discussion of results of spatial model

The spatial representation of monthly average PM_{10} and $PM_{2.5}$ concentrations in Delhi from January 2018 until March 2019 are depicted in Annexure I. The air pollution levels in Delhi show a strong seasonal variability with much lower pollution levels from July until September 2018 (monsoon season) and highest pollution levels in winter months of December and January. The spatial pattern of the concentration levels in Delhi is not significantly varying throughout the year, the absolute concentration levels do vary significantly. Each map shows the monthly average result from the RIO model and the monthly average station observations as overlay. Based on the results of the RIO model, it associates increased PM levels with more populated and more industrialized areas additional to the interpolation of the observations of the monitoring stations. The RIO model indicates several zones with increased PM levels, among which the following are identified as hotspots:

- ✓ Rampura Industrial Area and surroundings (28.680495◦N – 77.148103◦E)
- ✓ Narela Mandi – industrialized zone (28.831564◦N – 77.101166◦E)
- ✓ Bawana industrial area (28.789968◦N – 77.053402◦E)
- ✓ Rohini, Timarpur, North of Ashok Nagar
- ✓ Area around stations Jahangir Puri and Wazirpur
- ✓ Noida Sector 15 (28.586123◦ N – 77.313648◦ E)

An overlay of the annual average concentration maps for $PM_{2.5}$ and PM_{10} with the land use data has been made to calculate the landuse wise averages for both pollutants. Table 16 show the average concentration levels for the different land uses in Delhi (2018).

Table 16 Average concentration levels for different land uses in Delhi for 2018

Land use class	Total area (km^2)	$PM_{2.5}$ ($\mu g/m^3$)	SD $PM_{2.5}$ ($\mu g/m^3$)	PM_{10} ($\mu g/m^3$)	SD PM_{10} ($\mu g/m^3$)
Commercial Centre	13.5	116.8	±9.2	261.1	±22.0
Cultivated Land	551.2	105.9	±9.8	227.8	±14.8
Green Area	85.0	95.9	±18.8	219.5	±30.5
Industrial Area	28.3	129.9	±15.1	315.5	±40.1
Open Space	58.0	83.5	±16.7	199.7	±23.1
Others	40.6	83.7	±11.5	198.0	±13.8
Residential Area	680.9	115.0	±10.9	250.7	±21.1
Water Body	22.9	106.0	±13.6	231.5	±23.0
Grand Total	1481.4	108.4	±16.9	237.7	±34.9

Table 16 shows that the PM levels in different landuse categories of Delhi are varying considerably across different landuse categories. Annual average PM_{10} levels across different landuse categories in Delhi varied between 198 and 315.5 $\mu g/m^3$ with a standard deviation of ±34.9 $\mu g/m^3$ whereas the $PM_{2.5}$ levels ranged between 83.5 and 129.9 $\mu g/m^3$ with a standard deviation of ±16.9 $\mu g/m^3$. High PM levels are observed at industrial landuse category followed by commercial, residential landuse in Delhi. However, green area, cultivated land, water body etc. are also estimated with high PM levels in Delhi.

4.4 Conclusion: spatial modelling

- RIO has been set up successfully for the city of Delhi using data from all continuous stations operational in two time-frames a) 2015-2017, b) 2018 (Jan-Dec), c) 2019 (Jan-Mar)
- In this project, we focus on PM_{10} and $PM_{2.5}$ as the pollutant of concerns and pollutants like NO_2 and O_3 can be dealt later with the further studies on identification of appropriate drivers to account for local concentrations.
- Air pollution levels in Delhi show strong seasonal variability with much lower pollution levels from July until September 2018 (monsoon season) and highest pollution levels in winter months of December and January. The spatial pattern of the pollution remains fairly constant through the year. PM concentrations are highest in the northern part of Central Delhi and the densely populated area on the Yamuna river's east bank. The PM concentrations in New Delhi and the greener southern part and the city's outskirts (more cultivated land) are lower.
- A spatial driver beta has been developed for $PM_{2.5}$ based on land use due to significant correlations) with its initial conditions derived using the available emissions inventory. Correlation between PM concentrations and population density were found to be quite small and not used.
- Compared to other RIO set-ups in Belgium, Netherlands and Eastern Europe, the results for PM_{10} and $PM_{2.5}$ can be considered satisfactory.
- Another option to further improve RIO could be to use detailed traffic count information, if detailed traffic emissions are not available.
- Outliers (PM): Four stations do not agree well with the trends observed for other stations, mainly due to very local activities influencing their air quality patterns. These are Okhla Phase II, Najafgarh, Anand Vihar and Pusa road. They were dealt separately to refine the analysis.
- Looking at the land use foot prints, it is noteworthy to mention that the footprints of all stations are heavily affected by residential land use.
- RIO performed much better than the Ordinary Kriging and displays the local landuse influences.
- The RIO model successfully develops spatial maps for $PM_{2.5}$ and PM_{10} concentrations in Delhi, which can be used for hotspot planning and control.
- For both pollutants, O_3 and NO_2, a correlation analysis with several spatial drivers has been performed. No meaningful correlation could be found between NO_2 concentrations and most variables (population density, distance to roads, residential area fraction within buffer). Land use data did not prove to be a good spatial driver for NO_2 either. Further research would be needed to identify a suitable spatial driver for NO_2.
- The correlation analysis for O_3 showed more promising results. Interestingly, the residential land cover class shows quite a strong correlation across the different buffer radii which were tested, but the GHS population dataset did not show good correlations.
- A further research option would be to test satellite data with tropospheric ozone column densities and tropospheric NO_2 concentrations or high resolution emission maps for Delhi. Currently, we choose to focus on PM_{10} and $PM_{2.5}$ as these are the most important pollutants in the Delhi area.

5. OVL model: Forecasting setup

This chapter provides the details of the forecasting model setup for the two datasets – a) 2015-2017, b) 2018. The model forecasts are validated with the actual observations for three months in 2019 (Jan-Mar).

5.1 Forecasting setup (2015 – 2017)

The following sections describe the air quality forecasting model setup for the city of Delhi. For the purpose of training of the air quality forecasting model, the NCEP FNL (National Centre for Environmental Prediction Final) data historic archives have been used. Usage of ECMWF (European Centre for Medium-range Weather Forecast) data or local weather data would ideally be preferable as ECMWF performs better when compared with NCEP system (Roberto 2004); however FNL/GFS is freely available and well suited to test the methodology.

5.1.1 Methodology

Temporal and spatial variation of PM and other pollutants for a particular region is determined by complex interplay of many parameters. The primary objective is to develop a forecasting model for daily average PM and other pollutants at 0900 h of day 0. The model has been developed to predict the ground level values of pollutants like $<PM>_{day\ N}$,$<O_3>_{day\ N}$ and $<NO_2>_{day\ N}$ for days N=0,1,2,3,4, for different monitoring sites in Delhi. Special emphasis has been put on forecasting of typically high PM concentrations, which can then be used for triggering a warning signal. The forecasting model is setup based on the philosophy that forecasted pollutant concentrations will depend on previous days concentrations and forecasted meteorology for the next few days. The most important input parameters for the purposes of air pollutant forecasting are:

Initial conditions: variables measured before 9h of day0
Future conditions: forecasted meteorological variables
Target: $<pollutant>_{day\ N}$

For development of representative model, a large historical dataset of pollutant concentration for period a) 2015-2017 b) 2018 was collected and used. For each monitoring station, an artificial neural network NN was designed to establish a statistical relationship between chosen inputs and target forecasted pollutant values- $<PM>_{day\ N}$,$<O_3>_{day\ N}$ and $<NO_2>_{day\ N}$. For initial condition as input variables we considered: average pollutant concentration of first 9h of day0 and forecasted meteorological parameters includes: average boundary layer height (BLH), wind direction, temperature, RH, cloud cover etc. The interplay of different input parameter plays vital role in forecasting of air pollutants. Among these, a) BLH is a very important parameter as an input to NN, which determines the height up to which the pollutants can disperse due to turbulence in lower troposphere, b) wind speed determining the atmospheric advection of pollutants, c) temperature impacting the diffusion of pollutants, d) cloud cover influencing the formation of secondary pollutants, are the important ones consider for forecasting. Other than these, wind direction and day of week were the other two parameters considered as input variables since they can also play an important role in determining concentrations on a specific day or at a specific station. Moreover, for each monitoring station, ANN based forecasting model has been developed

and evaluations were carried out to assess the accuracy of forecasted values against actual observations.

5.1.1.1 Meteo-processing and correlation analysis

From the FNL data archives, we downloaded the meteorological data for 2015-2018 in grib format and extracted the following parameters (Table 17) for the Delhi study domain:

Table 17 Parameters extracted over the Delhi domain from FNL data archive

PAR	Meteo parameter
P01	2m temperature [K]
P02	2m relative humidity [%]
P03	10m v component of wind [m/s]
P04	10m u component of wind [m/s]
P05	planetary boundary layer height [M]
P06	Total cloud cover entire atmosphere
P07	high cloud cover
P08	low cloud cover
P09	medium cloud cover
P10	975mb - 1000mb inversion strength [K]
P11	950mb - 1000mb inversion strength (~500 m) [K]
P12	925mb - 1000mb inversion strength [K]
P13	900mb - 1000mb inversion strength (~1 km) [K]
P14	850mb - 1000mb inversion strength (~1.5 km) [K]
P15	dwsp/dz (shear stress) between bottom two layers (975mb - 1000mb)

More variables can also be included in this list depending on the availability of data (e.g. precipitation). First, a correlation analysis was performed between with the daily averaged concentrations for different stations and different variables. Based upon this analysis, parameters, which showed high correlations, were selected for construction of an initial model architecture to forecasts the day+N daily mean concentrations. The selected model parameters are presented in Table 18.

Table 18 Parameters for model input vector

Parameter	Description
CAVG_DAY0	Measured concentration average in the morning of day0, assuming the forecast will be started at 9h LT
CAVG_DAYM1	Measured concentration average for day-1
BLH_DAYN	Average dayN boundary layer height as calculated by the meteo model
T2M_DAYN	Average dayN 2m air temperature
RH_DAYN	Average dayN 2m relative humidity
IS975_DAYN	Average dayN temperature difference between 975 mb and 1000 mb model layer [K]
IS950_DAYN	Average dayN temperature difference between 950 mb and 1000 mb model layer [K]
IS925_DAYN	Average dayN temperature difference between 925 mb and 1000 mb model layer [K]
U10_DAYNM1_DAYN	Average zonal (west-east) wind speed between dayN-1 at 12:00 and dayN at12:00
V10_DAYNM1_DAYN	Average meridional (south-north) wind speed between dayN-1 at 12:00 and dayN at12:00
WEEKEND	0 : weekday, 1 : weekend

As an example correlation plots for the pollutant PM_{10} – day+1 forecasts is shown in Figure 52.

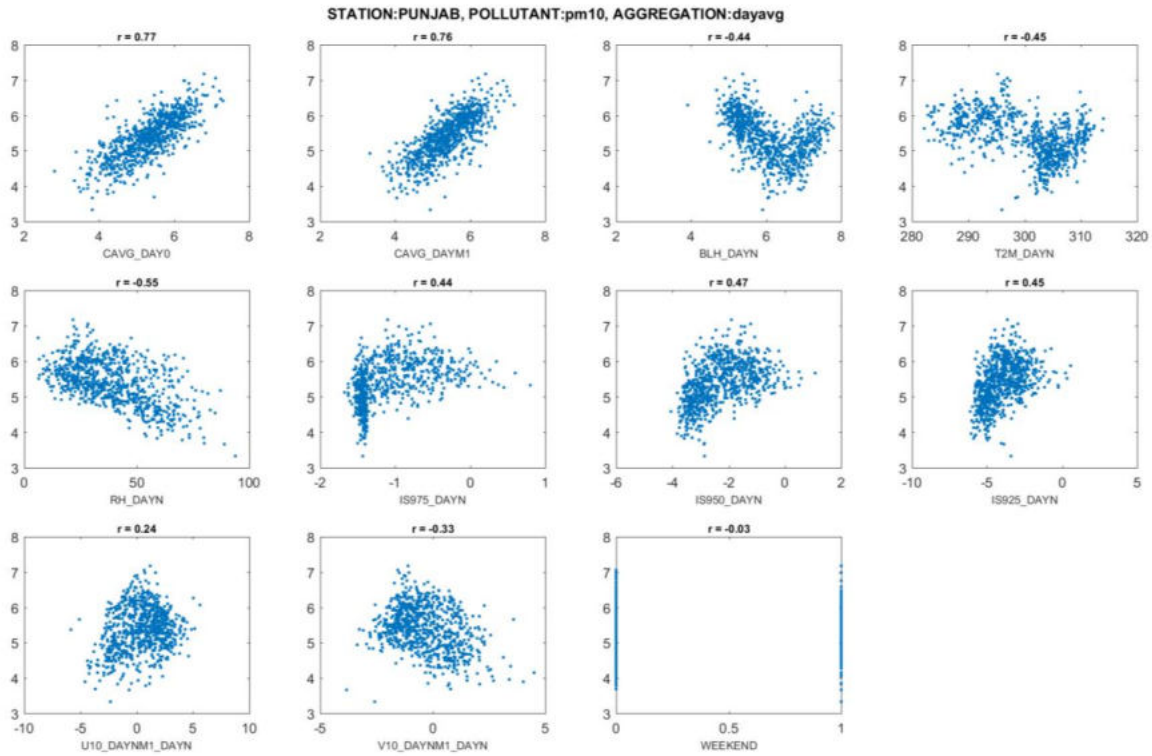


Figure 52 Correlation analysis for selected meteorological parameters in the ANN models. the horizontal axes show the input model parameter, the vertical axis shows the natural logarithm of the daily averaged PM_{10} concentrations.

As expected, high correlations were observed between PM_{10} concentrations with Day0 (same day concentrations till 9 AM) and Day M1 (previous day's concentrations). Interestingly the correlation with **Boundary Layer Height (BLH) shows a bi-modal structure**, which needs to be analyzed further. This could be due to the different seasons in India (wet vs. dry), circulation patterns, long range transport or mixing of particulate matter from aloft to ground levels. Inverse correlations have been observed between PM_{10} concentrations and RH and Temperature. Increased RH can lead to wet deposition, while higher temperature can also lead to lowering of PM_{10} concentrations due to more dispersion. Figure 53 shows a correlation analysis, averaged over different stations for each pollutant. Each bar represents the Pearson correlation coefficient for the day+ 1 daily averaged concentration with the Artificial Neural Network (ANN) model input vector.

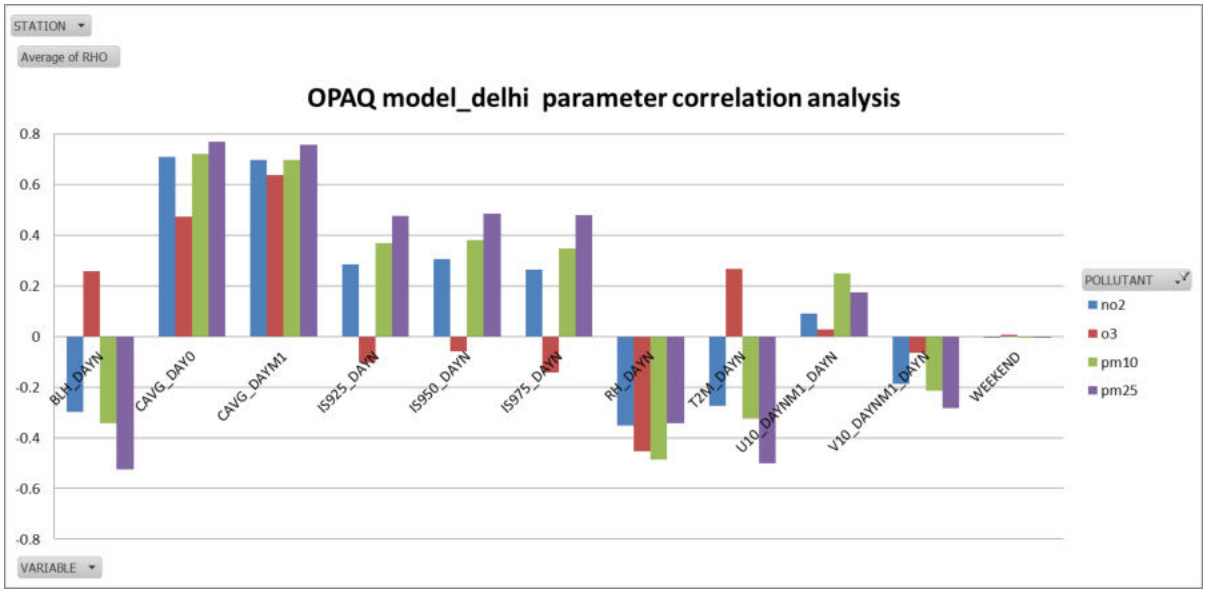


Figure 53 Correlation analysis, averaged over the stations for each pollutant

It is evident from figure 52 that:

a) correlation with the day 1 and day 0 observations is the strongest
b) O_3 shows opposite correlation with BLH than PM and NO_2, which is expected as ozone gets depleted due to titration reactions with local NO emissions.
c) RH shows negative correlation with all respective pollutants indicating downwash effect
d) Temperature (T2M) showed negative correlation with NO_2 and PM suggesting high PM concentration at low temperature. This is due to lesser dispersive conditions in low temperatures.

5.1.1.2 Neural network approach

For forecasting of air pollutants, we need to produce an output variable which should be a good estimate for the pollutant concentrations for the next (day1). This is to be based on a set of known input variables. This calls for a statistical model which fits the relationship between the input variables and targeted forecast of pollutant concentrations. The model hence needs to be developed based on historical datasets which depict relationships between input and output variables. These relationships are most likely non-linear, and hence basic parametric regression techniques are not so effective. In this case, a neural network approach is more appropriate. A detailed introduction to the NN techniques is given in Bishop (1995).

In this, for different monitoring site, a NN has to be designed to establish a fit function between the chosen inputs and targeted pollutant concentrations. Historical datasets are collected for both input and output variable. A part of the dataset is used for the training of the neural network and the rest is used for validation of the model.

5.2 Initial forecast validation results

The performance of NN and validation of forecast result is determined by comparing the measured pollutant concentration with forecasted precision value for each monitoring station. The measurement of performance is described in terms of various performance metrics- NMB, RMSE and R^2 between observed concentration value and forecasted result. NMB, RSME and R^2 denote Normalized mean bias, Root Mean Square Error value and coefficient of correlation. In this study, the forecast horizon is evaluated through the next 5 days for all monitoring stations.

Validation of results of different pollutants was carried out for various stations in Delhi. The day+1 forecast values are compared with actual observations. The normalized mean bias values between observed concentration and forecasts value is presented in Figure 54. It is evident that most of the stations show low negative bias values but some stations (like ANAD1 and PUNJAB) were observed with high biased value for O_3 and $PM_{2.5}$ indicating under estimation of model prediction.

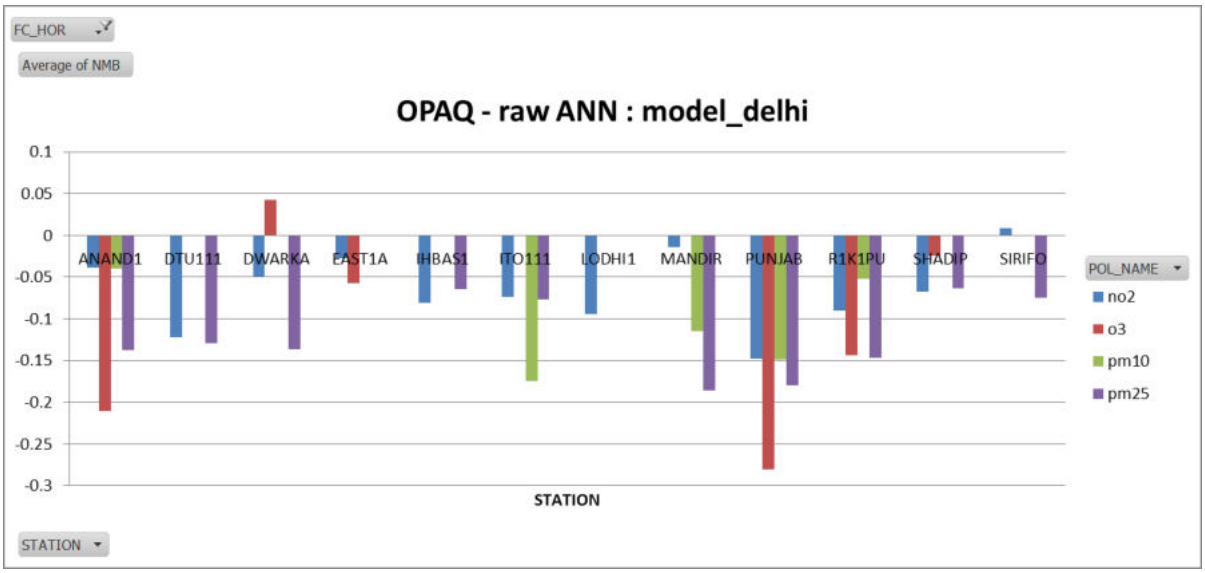


Figure 54 NMB value between observed concentration and forecasts for different pollutant at various stations (2017)

Figure 55 indicates RMSE value between observed concentration and forecasts. Evidently, most stations showed low error value, except at PUNJABI BAGH where RMSE was found to be higher for all pollutants. In terms of NO_2 , high error values were observed at ITO and LODHI road stations. Overall, the model has generated satisfactory forecasts for different pollutant concentration.

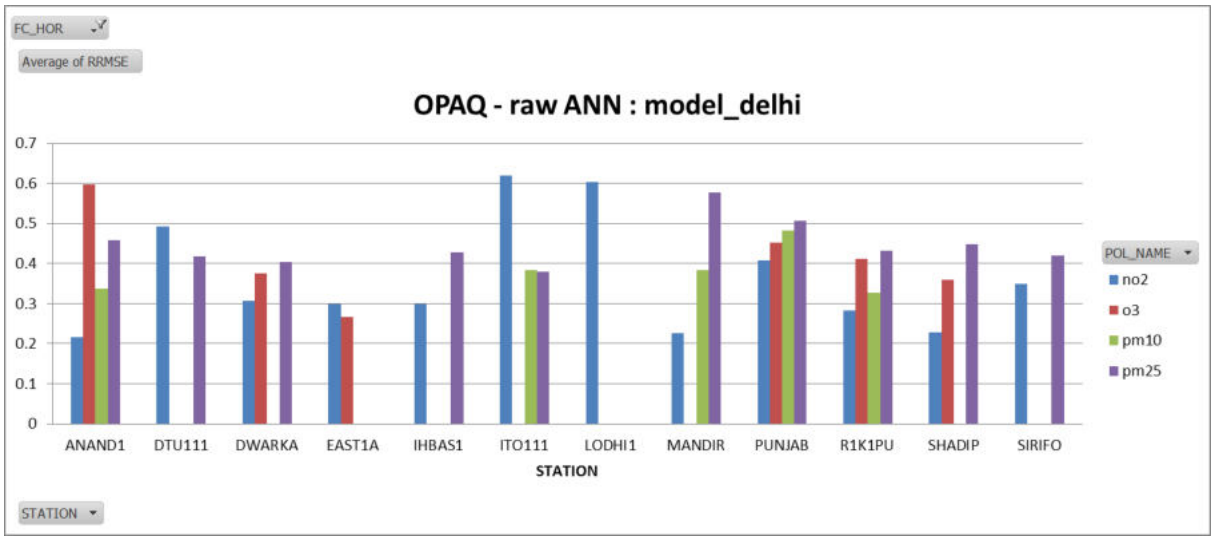


Figure 55 RMSE between observed concentration and forecasts for different stations and pollutants (2017)

Figure 56 indicates coefficient of correlation between forecasts and observed concentration for different pollutants at each of the monitoring location. NO_2 and $PM_{2.5}$ have shown higher correlation (R^2>0.5) for all the stations, whereas, PM_{10} was forecasted with high correlation at four out of five stations suggesting satisfactory performance of the forecasting model. At all the locations, O_3 shows poor correlation between the observed and the predicted values.

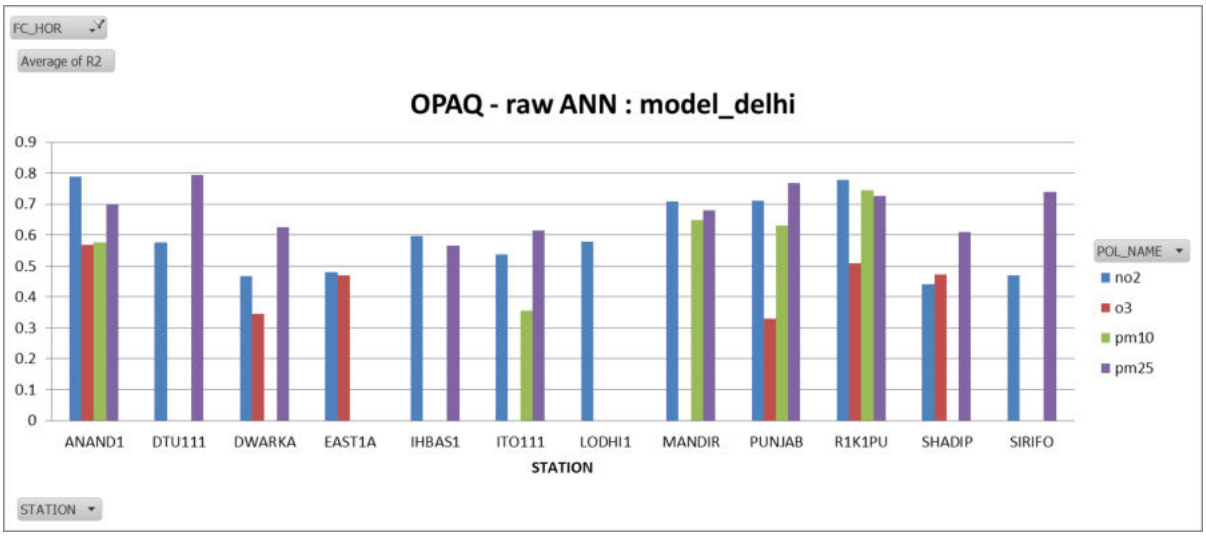


Figure 56 Coefficient of correlation between observed concentration and forecasted prediction value

5.2.1 Validation as a function of forecast horizon

Comparison of forecast model with observed concentration for forecast horizon of 5 days is also carried out. The performance of the model is again adjudged in terms of R^2 and NMB plots. Figure 57 and 58 show the temporal R^2 and NMB for daily average concentrations of all the stations for the 5-day forecast horizon.

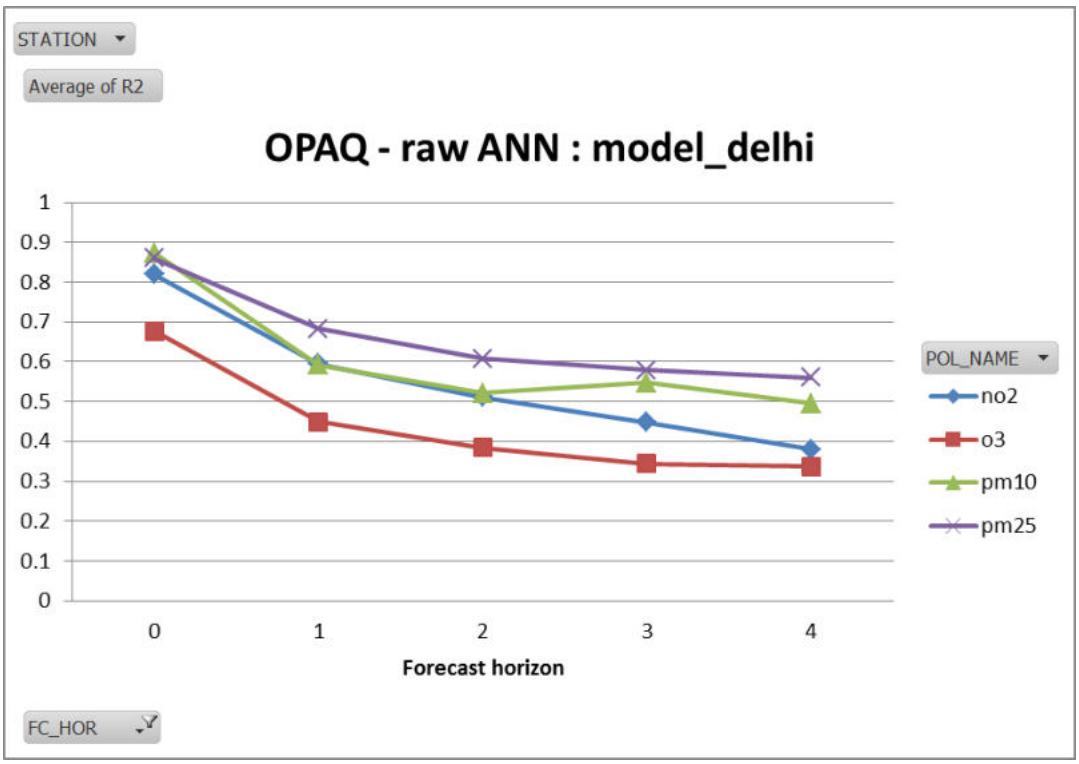


Figure 57 Coefficient of correlation between actual and predicted daily average concentrations in the forecast horizon of 5 days

*** Each dot represents an average over all the stations**

The model shows high correlation for day0 with R2>0.7 for all pollutants, however, the performance deteriorates over the next four days for all respective pollutant. Moreover, it can be interpreted that $PM_{2.5}$ showed better correlation throughout forecast horizon ti comparison to other pollutants- PM_{10}, NO_2 and O_3.

Figure 57 shows the normalized mean bias of the model for the forecast horizon of 5 days for different pollutants.

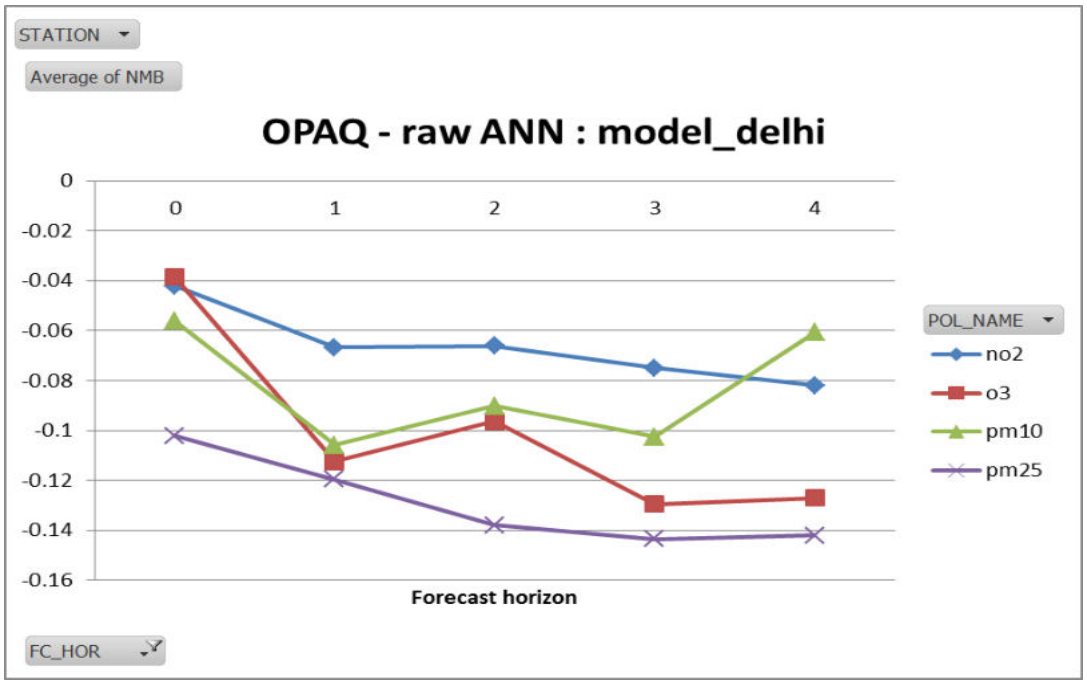


Figure 58 Normalized mean bias for daily average concentrations - raw ANN as a function of forecast horizon, each dot represents an average over the stations.

It is evident from figure 58 that the forecasting model is under-predicting and the bias was observed to be negative for all pollutants, throughout forecast horizon of four days. Minimum bias was witnessed for NO_2 whereas higher bias was observed for $PM_{2.5}$. Moreover, we can say model have satisfactorily forecasted concentrations of all pollutants showing limited bias for most locations indicating validation of forecasted results. The overall negative bias shows the under estimation of model prediction.

5.3 Forecasting model setup (2018)

Along with the preliminary findings for the monitoring data for 2015-2017, additional data of Jan-Dec 2018 was collected in order to improve the forecasting model. In addition to the parameters considered in the previous case, two new parameters have been added and analysed- Total precipitation from the FNL meteo archives and daily number of fire events in the surroundings of Delhi from the MODIS database.

5.3.1 Meteo processing and correlation analysis

From the FNL data archives, the data for 2018 has been also downloaded in grib format. Table 19 shows the parameters extracted over the Delhi domain from the FNL data archives.

Table 19 Parameters extracted over the Delhi domain from FNL data archive

PAR	Meteo parameter
P01	2m temperature [K]
P02	2m relative humidity [%]
P03	10m v component of wind [m/s]
P04	10m u component of wind [m/s]
P05	planetary boundary layer height [M]
P06	Total cloud cover entire atmosphere
P07	high cloud cover
P08	low cloud cover
P09	medium cloud cover
P10	975mb - 1000mb inversion strength [K]
P11	950mb - 1000mb inversion strength (~500 m) [K]
P12	925mb - 1000mb inversion strength [K]
P13	900mb - 1000mb inversion strength (~1 km) [K]
P14	850mb - 1000mb inversion strength (~1.5 km) [K]
P15	dwsp/dz (shear stress) between bottom two layers (975mb - 1000mb)
P16	Total precipitation

First, a correlation analysis was performed between the daily averaged pollutant concentrations for different stations and different parameters. Based upon this analysis, parameters showing high correlation were selected to formulate initial model architecture. The model was then used to forecasts the day+N daily mean concentrations. Table 20 represents the parameters selected for model input vector.

Table 20 Parameters for the model input vector

Parameter	Description
CAVG_DAY0	Measured concentration average in the morning of day0, assuming the forecast will be started at 9h LT
CAVG_DAYM1	Measured concentration average for day-1
BLH_DAYN	Average dayN boundary layer height as calculated by the meteo model
T2M_DAYN	Average dayN 2m air temperature
RH_DAYN	Average dayN 2m relative humidity
IS975_DAYN	Average dayN temperature difference between 975 mb and 1000 mb model layer [K]
IS950_DAYN	Average dayN temperature difference between 950 mb and 1000 mb model layer [K]
IS925_DAYN	Average dayN temperature difference between 925 mb and 1000 mb model layer [K]
U10_DAYNM1_DAYN	Average zonal (west-east) wind speed between dayN-1 at 12:00 and dayN at12:00
V10_DAYNM1_DAYN	Average meridional (south-north) wind speed between dayN-1 at 12:00 and dayN at12:00
TP_DAYN	Average dayN total precipitation
FIRE_DAYN	Average number of fire events during three day period M1-0-N (day – 1, current day, day+1), threshold of 10 events for this average.

Compared to the setup for 2015-2017, the parameter weekend has been dropped as no significant correlation between this parameter and observations is found. The number of fire events has been averaged over a three days period and a threshold value of 10 events has been applied to remove noise. As an example, correlation plots for PM_{10} – day+1 forecasts for a station (Wazirpur) with different parameters are shown in Figure 59.

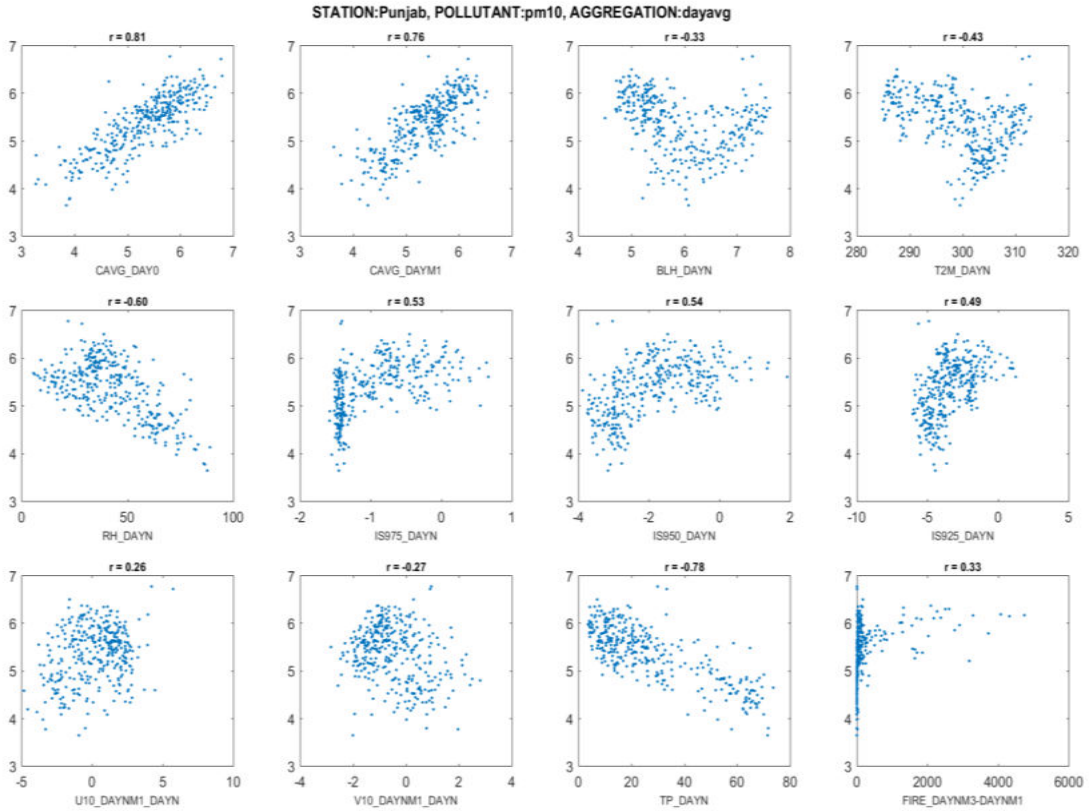


Figure 59 Correlation analysis for selected meteo parameters, concentrations and fire events in the ANN models.

* The horizontal axes show the input model parameter, the vertical axis shows the natural logarithm of the daily averaged PM_{10} concentrations.

The following important findings were observed:

- Strong correlations with concentrations on day0 and day-1.
- Interesting bi-modal structure in the correlation with boundary-layer-height
- Strong inverse correlation with total precipitations
- Limited correlation with number of fire events

Figure 60 shows correlation analysis averaged over the stations for various pollutants. Each bar represents the Pearson correlation coefficient for the day+ 1 daily averaged concentration with the ANN model input vector.

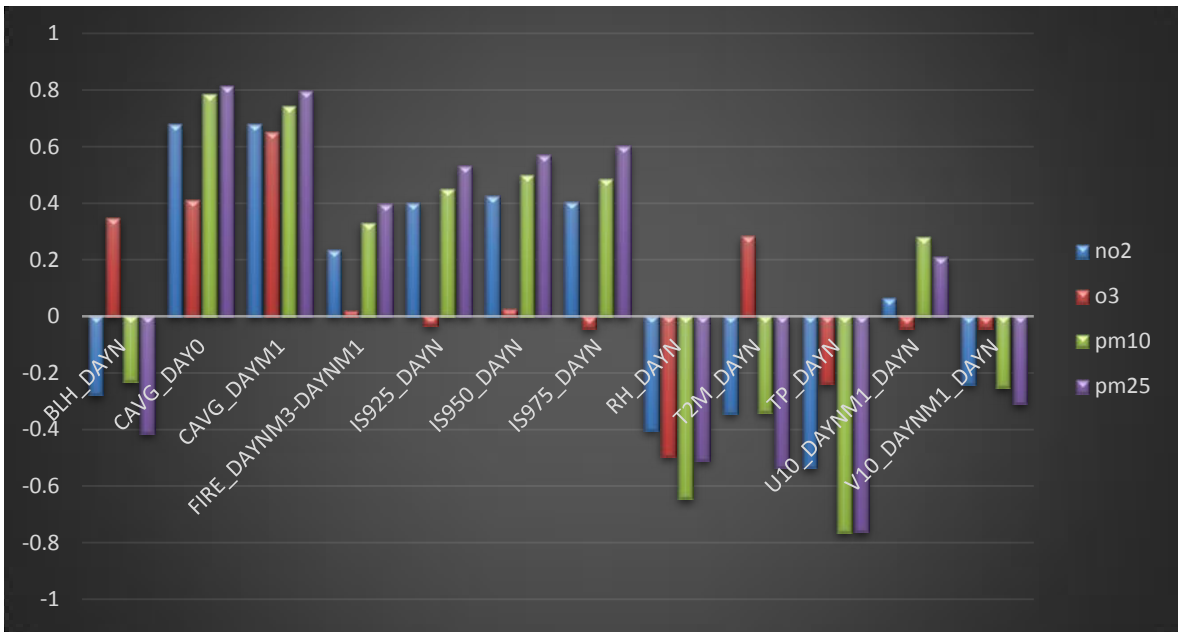


Figure 60 Correlation analysis averaged over all stations for day+1 averaged concentrations based on observations from 01/01/2018 – 31/12/2018.

This can be observed from figure 60 that

- Correlations were strongest for day-1 and day0 observations and total precipitation
- O_3 shows opposite correlation with BLH and temperature than PM and NO_2, which is normal as O_3 gets reacted away due to local (NOx) emissions during stable episodes.
- Correlation with number of fire events is fairly limited.

The limited correlation with the number of fire events is further investigated. Figure 61 shows the number of fire events from January 2017 until March 2019. Fire events are a yearly returning phenomenon around Delhi with a large peak in summers (burning wheat crop residues) and post monsoon seasons (due to burning of rice crop residues). The correlation with the number of fire events is fairly limited. Only during after the harvesting season, the number of fire events increases significantly and during the rest of the year this parameters remains low and uncorrelated with the changes in concentration levels.

This is to be noted that the impact of these fire events will not only depend on the number of events and time of the year, but also on the dominant wind direction.

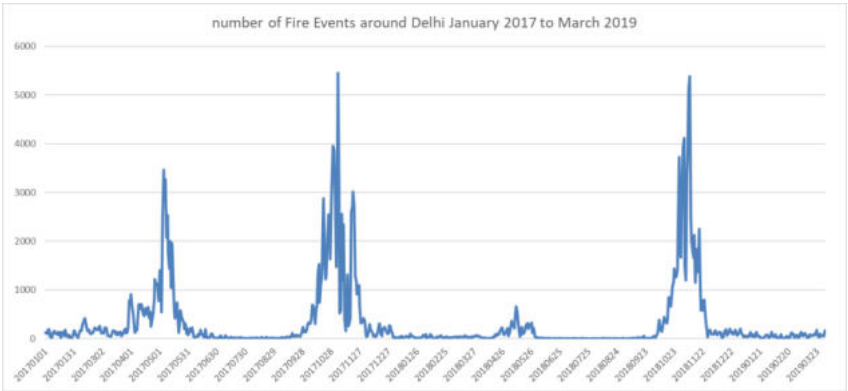


Figure 61 Daily number of fire events around Delhi from January 2017 until March 2019

5.3.2 Forecast model setup 2018

A model with the twelve parameters listed in the previous section has been trained and optimized for $PM_{2.5}$, PM_{10}, O_3 and NO_2: CAVG_DAY0, CAVG_DAYM1, BLH_DAYN, T2M_DAYN, RH_DAYN, IS975_DAYN, IS950_DAYN, IS925_DAYN, U10_DAYNM1_DAYN, V10_DAYNM1_DAYN, TP_DAYN, FIRE_DAYN. This model has been applied, trained, optimized and validated for four different types of artificial neural networks (ANN):

- No real time corrections (RTC) - No resampling (RS) (0-0), raw ANN
- No real time corrections - Resampling (0-1)
- Real time corrections - No resampling (1-0)
- Real time corrections - Resampling (1-1)

Real-time corrections: Based on the difference between the forecasted values and the observations during the past days, a correction can be applied to the forecast values. This is called real-time-corrections (or RTC). During the optimization of a neural network using RTC, the number of days on which to base the corrections is optimized using the historical data. The number of days varies between 0 (no RTC) and 20 days of observations on which to base the correction. The advantage of adding RTC, is that it enables to almost completely remove the bias in the forecasted values. The disadvantage might be that the forecast respond too slow to a change in conditions.
In addition to the raw neural network output, the OPAQ forecast module contains a dynamic bias correction scheme as well. The daily forecasts as well as the observed values (as they become available) are stored in a temporary database. From this database a timeseries with past forecast errors can be reconstructed. Via a number of different methods a dynamic estimation of the current forecast error can be made. This can either simply be an average of the past forecast errors for a given forecast horizon, or a weighted average, where the errors further in the past are given less weight in the estimation of the current forecast error. When applying the estimated forecast error to the raw output of the ANN models, a "real time corrected" forecast is constructed. Again, the hindcast period over which to consider the forecast errors as well as the method and parameters for the estimation of the current forecast errors are optimized per station and per forecast horizon.

Resampling: Resampling is used to ensure a dataset has sufficient observations with increased concentrations. Instead of using a dataset 'as is', the dataset can be resampled for training and optimization, so it can be better trained for these occasions with increased concentrations.

5.3.4 Forecast model validation results (2018)

A comparison of the model results for all pollutants is made using different performance metrics e.g. R^2, RMSE, and bias. The sensitivity of model outputs was also tested with and without RTC and RS options. The results for $PM_{2.5}$ have been used to assess the impact of RTC and RS. All other pollutants have also shown similar behavior.
The sensitivity assessment shows that real-time-corrections have a positive impact on the model results, as it leads to lower relative RMSE and corrects for most of the bias. However, resampling has a negative impact on the validation, leading to higher relative RMSE, lower

R^2 and slightly increased normalized mean bias. This is not surprising as the observations in Delhi have sufficient periods with increased concentrations and resampling may not add further value. We therefore use real-time-corrections and do not include resampling for all four pollutants.

Figure 61 shows the relative RMSE averaged over all stations as a function of forecast horizon from day0 to day4 for $PM_{2.5}$. It is evident from Figure 62 that RMSE was lowest for the model optimized with real time correction (RTC) and No Resampling (RS) throughout the forecast horizon of day0 to day4. The model was trained and optimized with least error on day0, however for subsequent days, higher error was observed with No RTC and RS for PM_{10}, O_3 and NO_2. In case of $PM_{2.5}$, high error was accounted for model trained with No RTC and No RS optimization. The model validation was found to be best with the RTC and without RS options. Overall, this can be concluded that the model shows satisfactory performance in forecasting the concentrations of various pollutants./

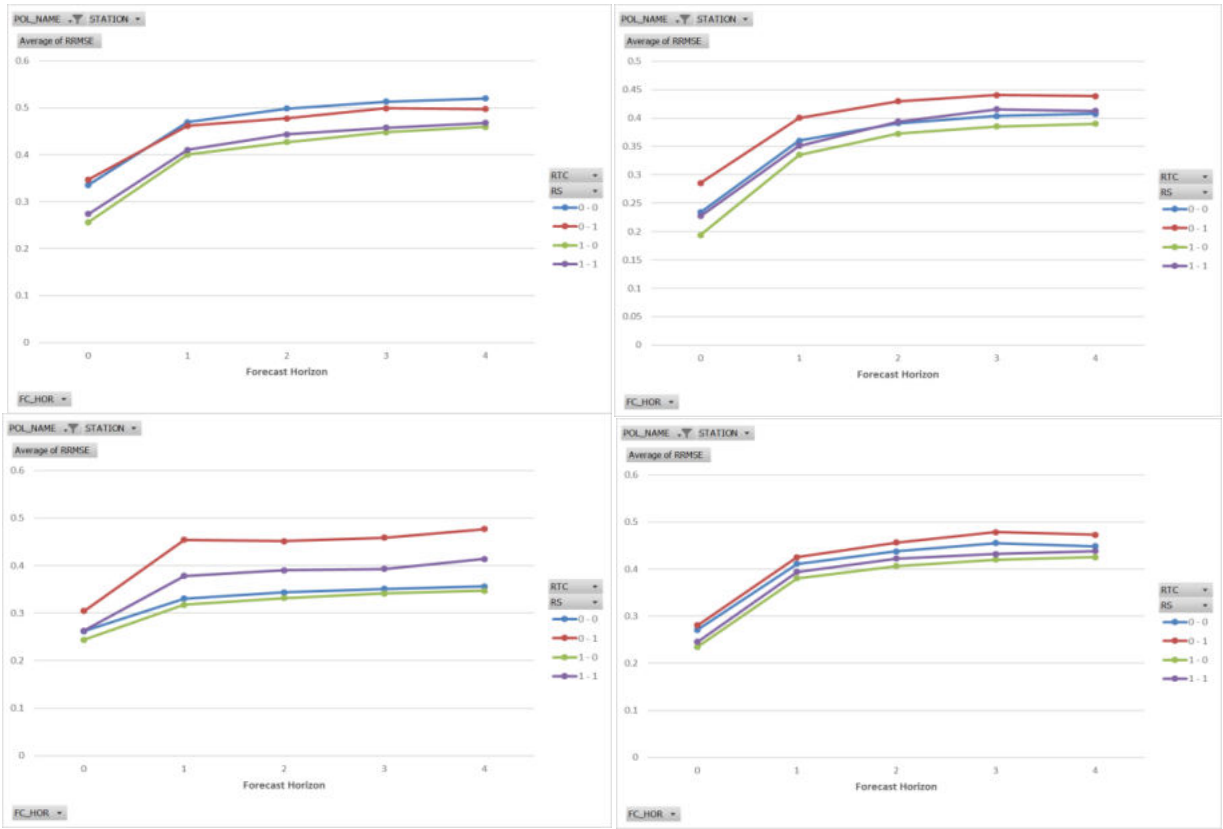


Figure 62 Comparison of relative RMSE averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2.
Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS.

Figure 63 represents average relative RMSE for each station considering forecast horizon is made for day+1 i.e. next day for all pollutants. The model was forecasted with relative high error >0.5 for NO_2 at 3 stations however for most locations relative RMSE was less varying between 0.3-0.4 indicating good validation of forecast output for all pollutants.

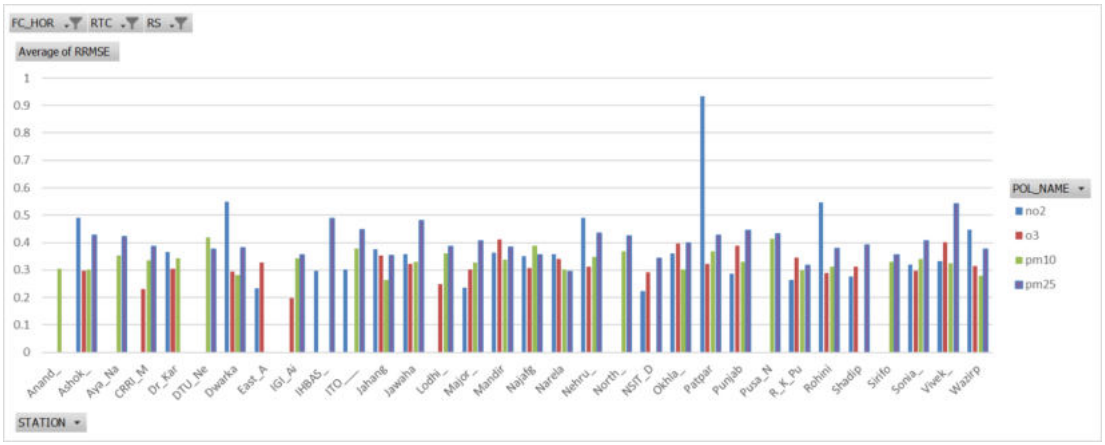


Figure 63 Relative RMSE per station for OPAQ Delhi model for year 2018 with RTC, no Resampling, forecast horizon day+1. Blue: NO2; Red: O3; Green: PM10; Purple: PM2.5

Figure 64 represents the comparison of normalized mean bias averaged over all stations as function of forecast horizon of day0 to day4 for all pollutants. The model shows low bias for all the pollutants with RTC-No RS settings, whereas considering other optimization scenarios the model forecasted negative bias indicating under estimation of forecast result. The RTC-RS forecast scenarios have also shown low bias suggesting good validation of model output. Overall, the NMB was found to be very low in all the forecasting scenarios.

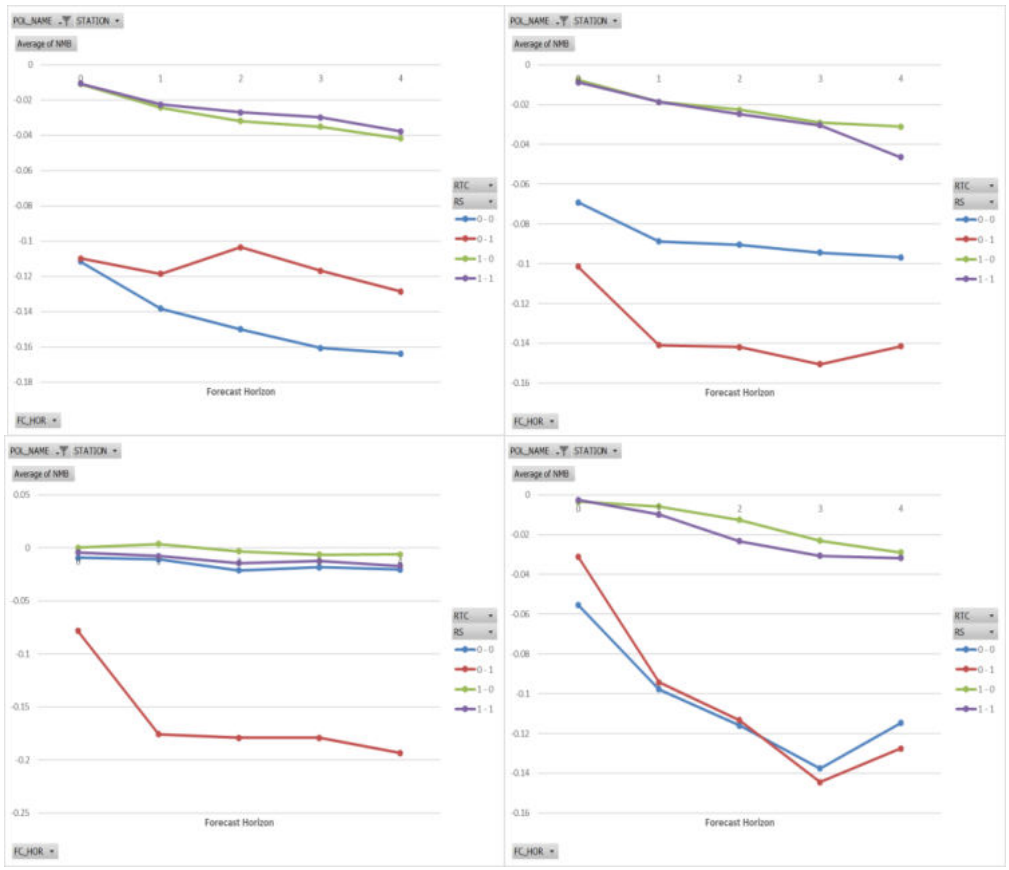


Figure 64 Comparison of normalized mean bias averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2. Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS.

Figure 65 indicates average normalized mean bias between observed concentration and forecasted value for each station for the forecast horizon of day+1 i.e. next day for respective pollutants. The model forecast showed less bias for most stations. However two stations, Najafgarh and Wazirpur were predicted with higher negative and positive bias for PM_{10} and O_3, respectively, suggesting under estimation and over estimation of forecast output result.

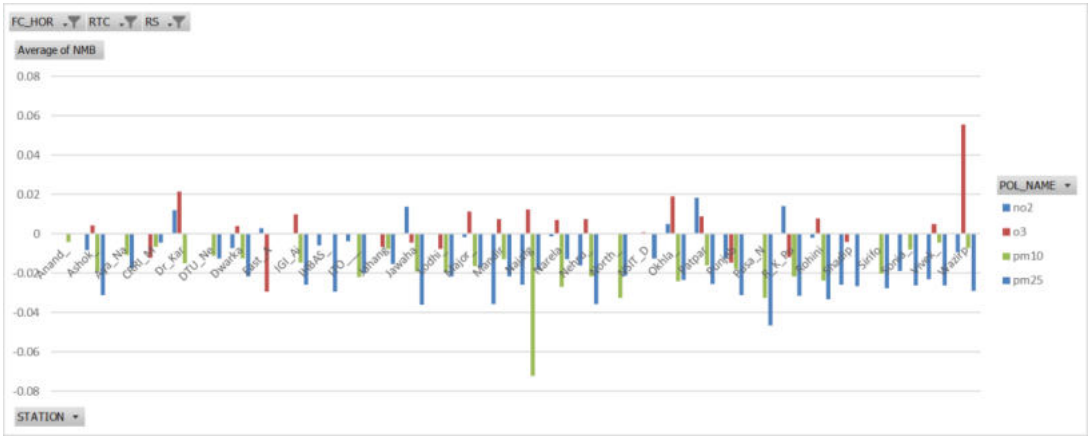


Figure 65 Average normalized mean bias per station for OPAQ Delhi model for year 2018 with RTC, no Resampling, forecast horizon day+1. Blue: NO2; Red: O3; Green: PM10; Purple: PM2.5.

Figure 66 represents the coefficient of correlation between the observed concentrations and forecasted results (averaged over all stations) for the forecast horizon of day0 to day4. It is evident that correlation was high for model trained and optimized considering No RTC-No RS and RTC-No RS event. The high correlation with R^2>0.8 was observed for model trained and optimized with No RTC and RS for day 0, and it deteriorates in subsequent days.

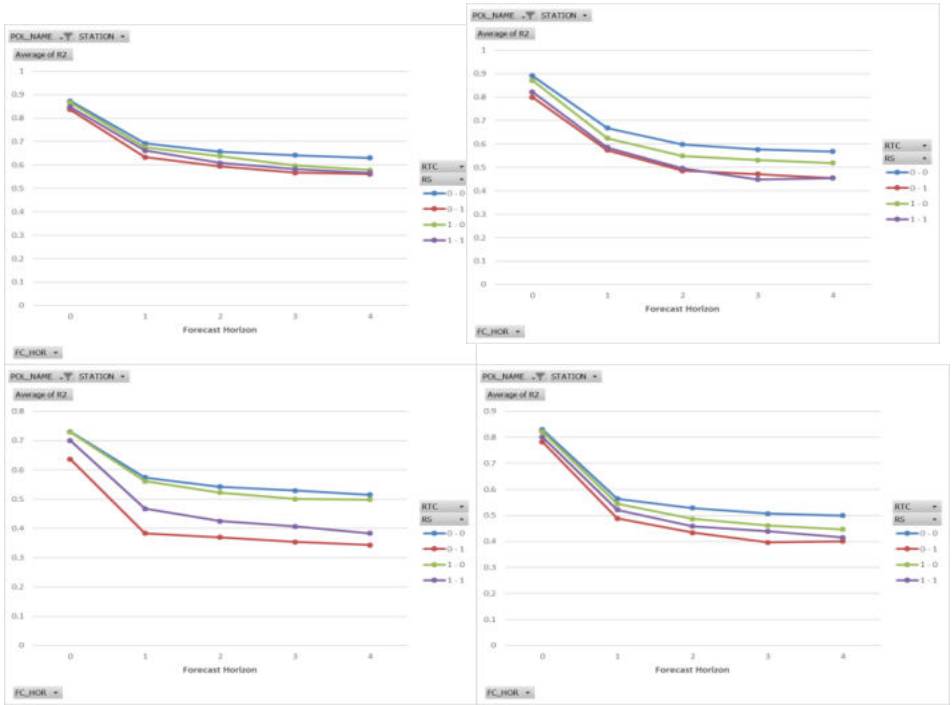


Figure 66 Comparison of R2 averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3/ and (d) NO2. Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS

Figure 67 indicates average relative R2 between observed concentration and forecasted value for each station considering forecast horizon is made for day+1 i.e. next day for respective pollutants. The model forecast showed good correlation with observed concentration having R2>0.5 for most stations for all pollutants indicating good validation of forecast output. However, for NO2 for few stations (e.g. Dwarka), the model results showed less correlation with observed concentration

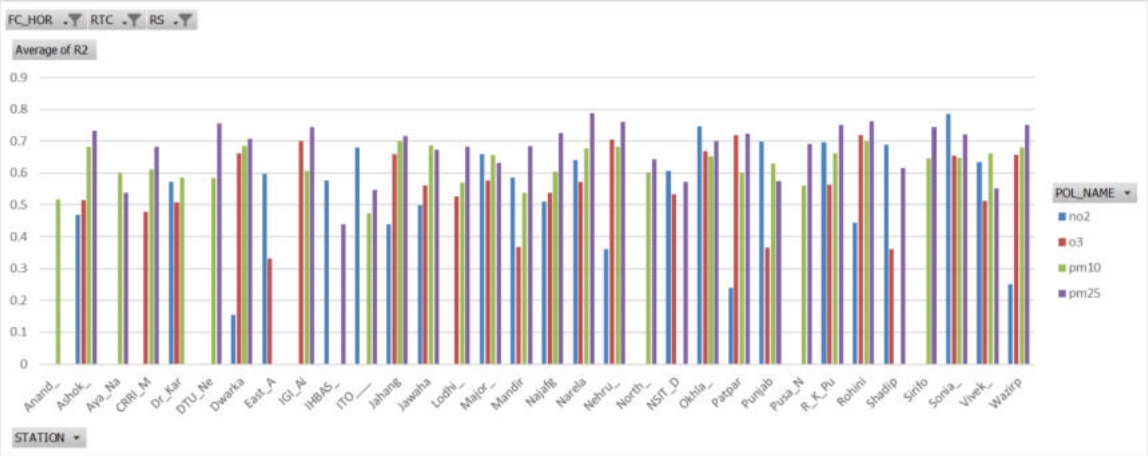


Figure 67 Average R2 per station for OPAQ Delhi model for year 2018 with RTC, no Resampling, forecast horizon day+1. Blue: NO2; Red: O3; Green: PM10; Purple: PM2.5.

Figure 68 represents the fraction of correct forecast (FCF) between the observed concentrations and forecasted results (averaged over all stations) for the forecast horizon of day0 to day4 for pollutant $PM_{2.5}$ and PM_{10} respectively. It is evident that FCF was high for model trained and optimized considering RTC-No RS and RTC-RS options. The FCF was more than 80 for day0, however, went down for subsequent days. Better forecasts were obtained for model trained and optimized with RTC and RS options. For PM_{10} forecasts were somewhat better than $PM_{10}$for various options.

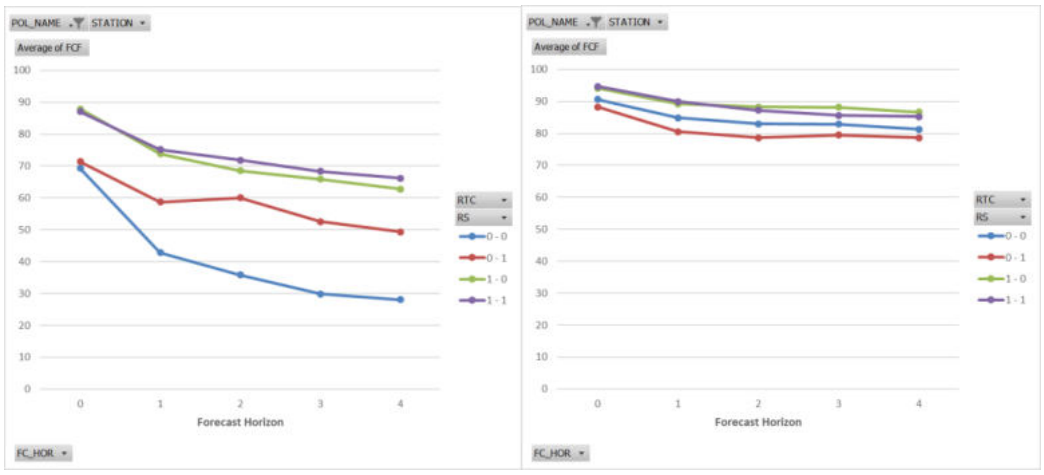


Figure 68 Comparison of FCF averaged over all stations as a function of forecast horizon for (a) PM2.5 and (b) PM10. Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS.

Figure 69 indicates average FCF between observed concentration and forecasted value for each station considering forecast horizon is made for day+1 i.e. next day for PM_{10} and $PM_{2.5}$ respectively. The model forecast showed FCF more than 60 for most stations indicating good validation of forecast output. However, in case of $PM_{2.5}$ for few stations (e.g. IHBAS), the model results showed less FCF (<60) with observed concentration.

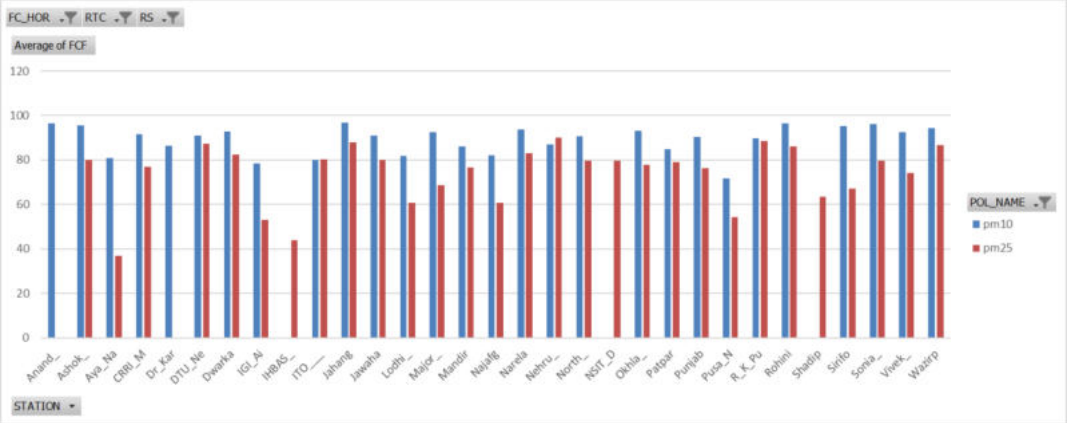


Figure 69 Average FCF of stations in Delhi for year 2018 for forecast horizon day+1. Blue: PM10; Red: PM2.5

Figure 70 represents the Fraction of false alert (FFA) between the observed concentrations and forecasted result (average over all stations) for the forecast horizon of day0 to day4 for pollutant $PM_{2.5}$ and PM_{10} respectively. It is evident that FFA was less for model trained and optimized considering No RTC-No RS event. The least false alert with FFA<5 was forecasted during day0. Number of events with high false alert was obtained for model trained and optimized with RTC and RS for both $PM_{2.5}$ and PM_{10} respectively.

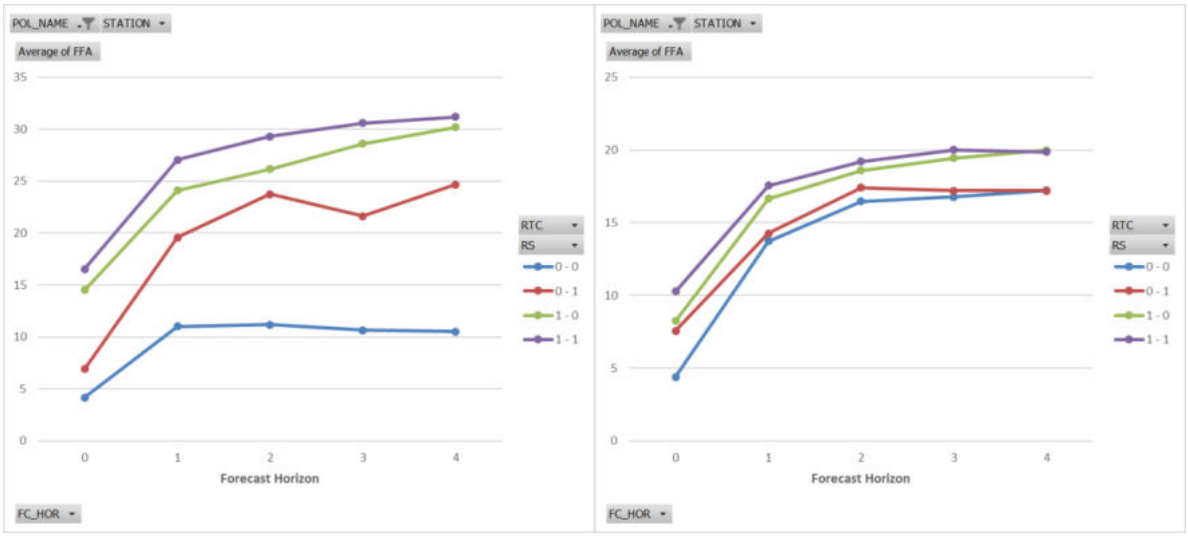


Figure 70 Comparison of FFA averaged over all stations as a function of forecast horizon for (a) $PM_{2.5}$ and (b) PM_{10}. *Blue: no RTC – no RS; Red: no RTC – RS; Green: RTC – no RS; Purple: RTC – RS*

Figure 71 indicates average FFA between observed concentration and forecasted value for each station considering forecast horizon is made for day+1 i.e. next day for PM_{10} and $PM_{2.5}$ respectively. The model forecast showed less false alerts with observed concentration having FFA<15 for most stations indicating good validation of forecasted output. However, for few stations (e.g. Aya Nagar), the model results showed high false alerts having FFA>25 with observed concentration.

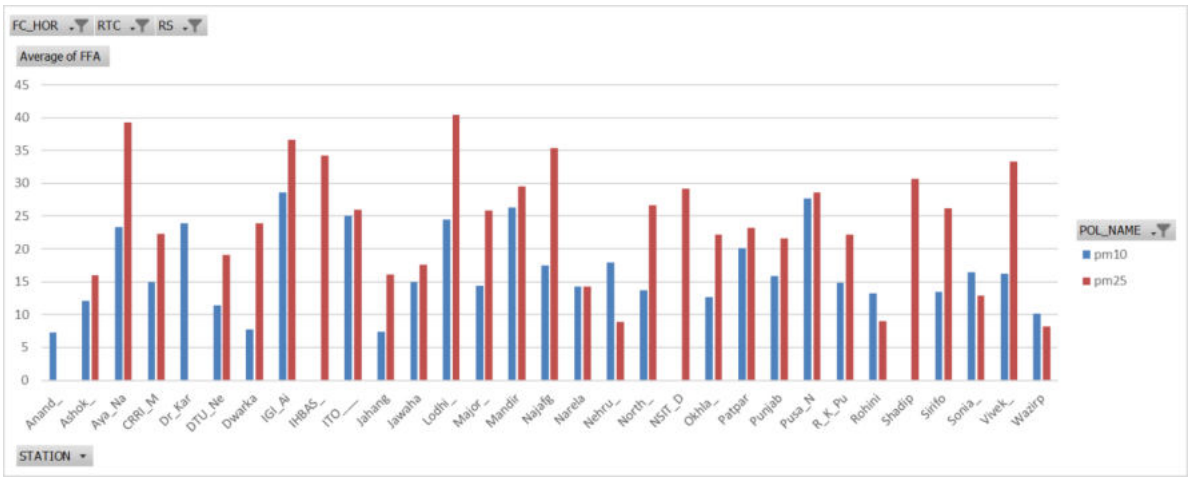


Figure 71 Average FFA per station for OPAQ Delhi model for year 2018 for forecast horizon day+1. Blue: PM10; Red: PM2.5

5.4 Forecasting model pilot testing for 2019

The OPAQ forecasting chain has been trained and optimized for 2018. The model has been tested for three months (Jan-Mar) in 2019. For purpose of research, three different forecasting models (based on selection of variables) are tested:

Model1: 11 variables:
CAVG_DAY0, CAVG_DAYM1, BLH_DAYN, T2M_DAYN, RH_DAYN, IS975_DAYN, IS950_DAYN, IS925_DAYN, U10_DAYNM1_DAYN, V10_DAYNM1_DAYN, TP_DAYN

Model2: 12 variables:
Model1 + FIRE_DAYNM3-DAYNM1 (average number of fire events in the last three days prior to the forecast day).

Model3: 4 variables:
CAVG_DAY0, BLH_DAYN, TP_DAYN, FIRE_DAYNM3-DAYNM1

The observations for 2019 have been downloaded and converted to the OPAQ input format. The necessary meteo data have been downloaded as well (NCEP FNL - National Centre for Environmental Prediction Final).

As first check, the overview of the available stations in 2019 has been made:

Table 25 Number of stations with sufficient data in 2018 and 2019

	2018	2019	Comments
PM_{10}	28	28	None
$PM_{2.5}$	29	29	Two additional stations for 2019 (not used: Anand_ and Dr_Kar)
NO_2	23	16	7 stations had NO_2 data in 2018 but lack for 2019, 9 stations have sufficient data in 2019 but could not be trained on 2018 due to insufficient data
O_3	24	24	7 additional stations available for 2019 which have not been trained on 2018 due to insufficient data

Validation statistics 2019 are given in the figures below for all 4 pollutants as a function of forecast horizon. All statistics are averaged over all stations in Delhi (with sufficient observations, see table above). OPAQ is configured including real-time corrections.

All three models have been tested for the first 3 months of 2019 on **daily average** basis. Based on the findings of 2018 dataset, all three models are tested with real-time-corrections and without re-sampling options. To run the validation, the observations for the first three months of 2019 have been converted to the input format of OPAQ. The FNL meteo data have been downloaded and converted to the input format.A comparison of all model results for all pollutants is made using different performance metrics e.g. R^2, RMSE, and bias.

Figure 72 shows the relative RMSE averaged over all stations as a function of forecast horizon from day0 to day4 for different pollutants tested on three different model run scenarios. It is evident that RMSE was lowest for model1 and model2 throughout the forecast horizon of day0 to day4. During the validation of model1 and model2, the forecasted values showed least error on day0 with RMSE<0.3 however, considering model3, the forecasted value was observed with higher error throughout the forecast horizon for all

the pollutants. Overall, the output suggests adding the variable of number of fire events to model1, model2 does not improve the prediction of the concentrations.

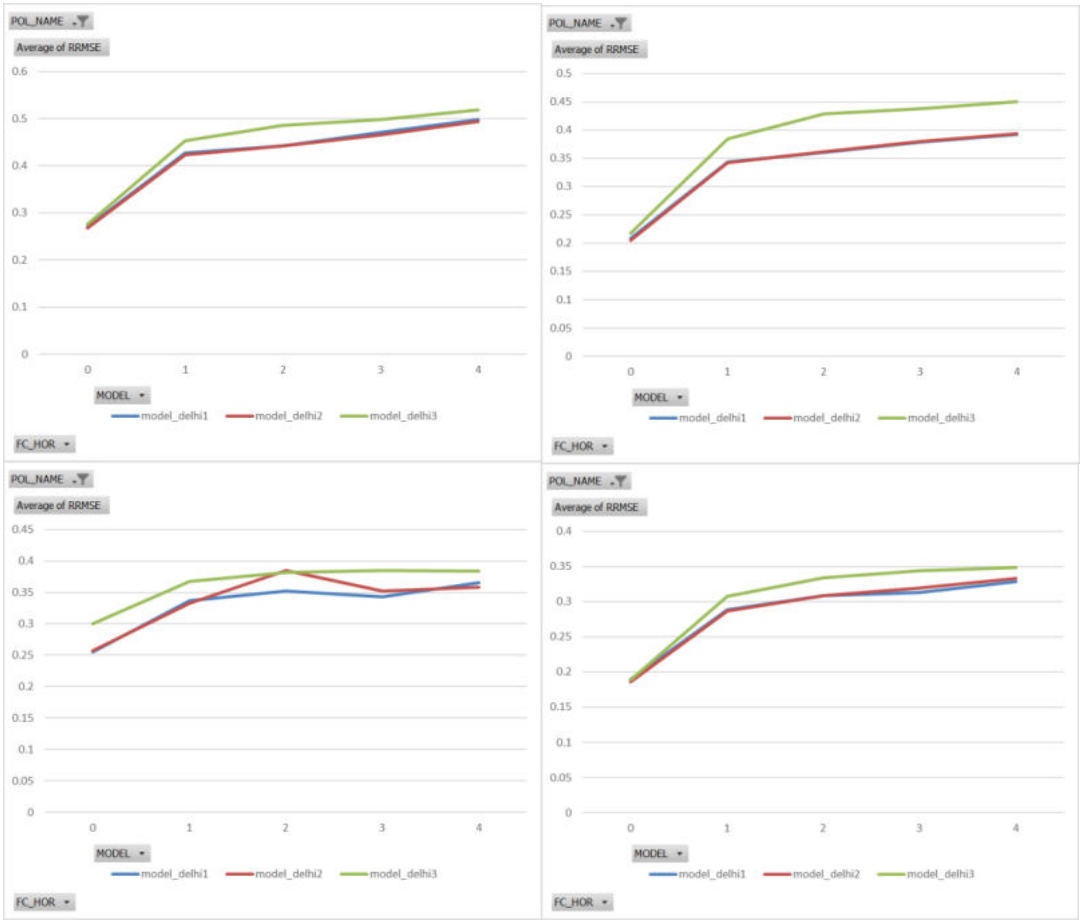


Figure 72 Comparison of relative RMSE considering all the models averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2.

Figure 73 shows the comparison of normalized mean bias averaged over all stations as function of forecast horizon of day0 to day4 for different pollutants based on three different model run scenarios. The model1 showed low bias for all the pollutants compared to model2 and model3, whereas considering other model run scenarios the forecasted values showed more positive and negative bias indicating over and under estimation of forecasted output. It was evident that adding the number of fire events data to model2 does not show improved output compared to model1. Considering model3, the optimization was observed with high bias in predicting the concentration. Overall, under model1 optimization, the NMB was found to be very low for all respective pollutants, hence, can be concluded as the best model in the present circumstances

Figure 73 Comparison of normalized mean bias considering all models averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2

Figure 74 shows the coefficient of correlation between the observed concentrations and forecasted result (averaged over all stations) for the forecast horizon of day0 to day4 for all pollutants based on three different model run scenarios. It is evident that correlation was high for model trained and optimized with model1 and model2 compared to model3. The high correlation with $R^2>0.7$ was forecasted during day0 for pollutant $PM_{2.5}$, PM_{10} and NO_2 whereas $R^2<0.6$ was forecasted for O_3. Overall, the less correlation was forecasted for model trained and optimized with model3. It was also evident considering number of fire events data in model2 did not improve forecasting results of the pollutant concentration. Moreover, it can be interpreted that both model1 and model2 showed satisfactory performance in predicting the concentration for respective pollutants.

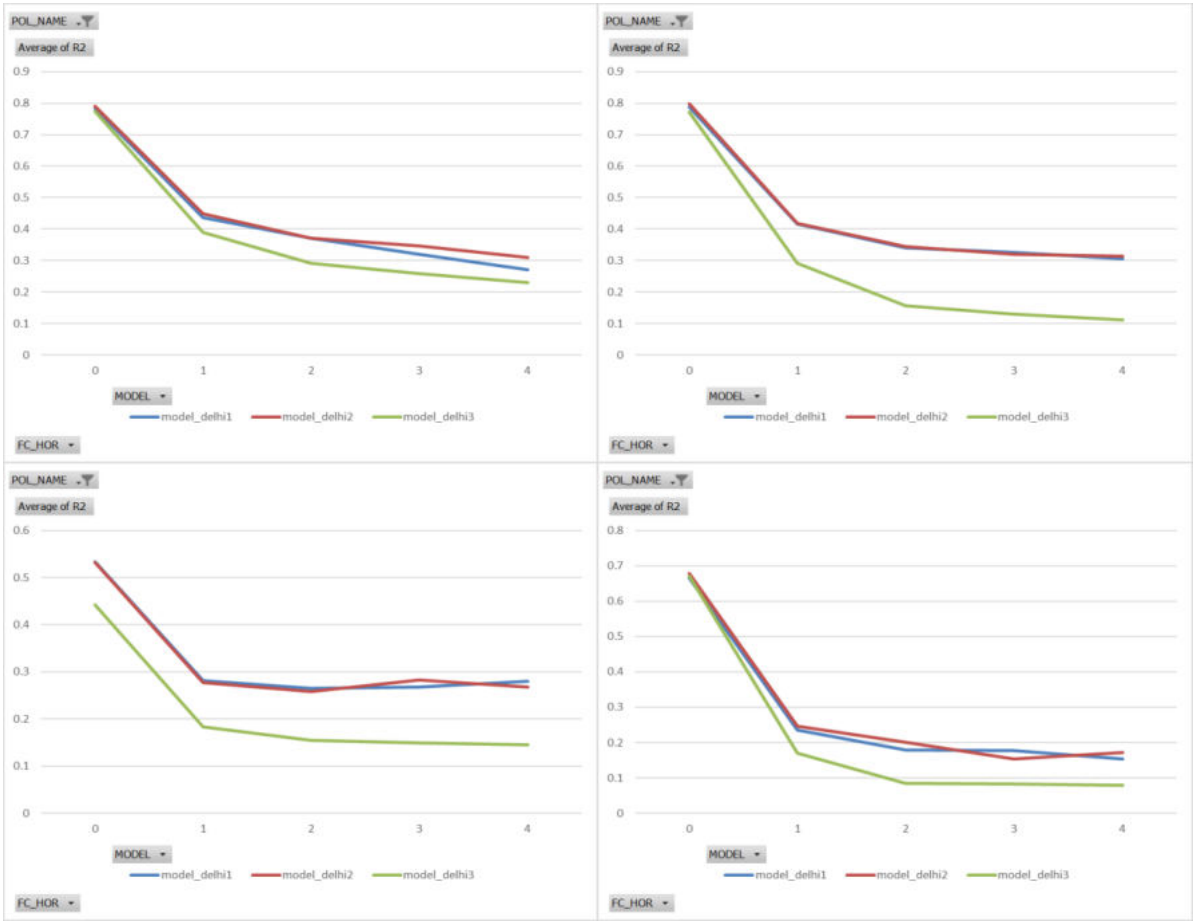


Figure 74 Comparison of R2 considering all models averaged over all stations for the forecast horizon for (a) PM2.5, (b) PM10, (c) O3 and (d) NO2

5.5 Time series analysis of forecasts

Time-series plots were prepared for several stations to compare forecasts with actual observations. Figure 75 presents the daily observed average $PM_{2.5}$ concentrations at Ashok Vihar station and its comparison with forecasted concentrations in different scenario: forecast concentration for day+1 with no RTC and forecast concentration for day+1 with RTC. It could be interpreted that the forecasted concentration with RTC was close to observation value for most of the time throughout monitoring period. For both the forecast scenarios, the predicted concentrations were close to the observed concentrations for all months except for a few observations in the initial months and some extremely high concentrations in between. Very high concentrations are observed in the months of October to December, mainly on account of agricultural residue burning events and metereological adversity.

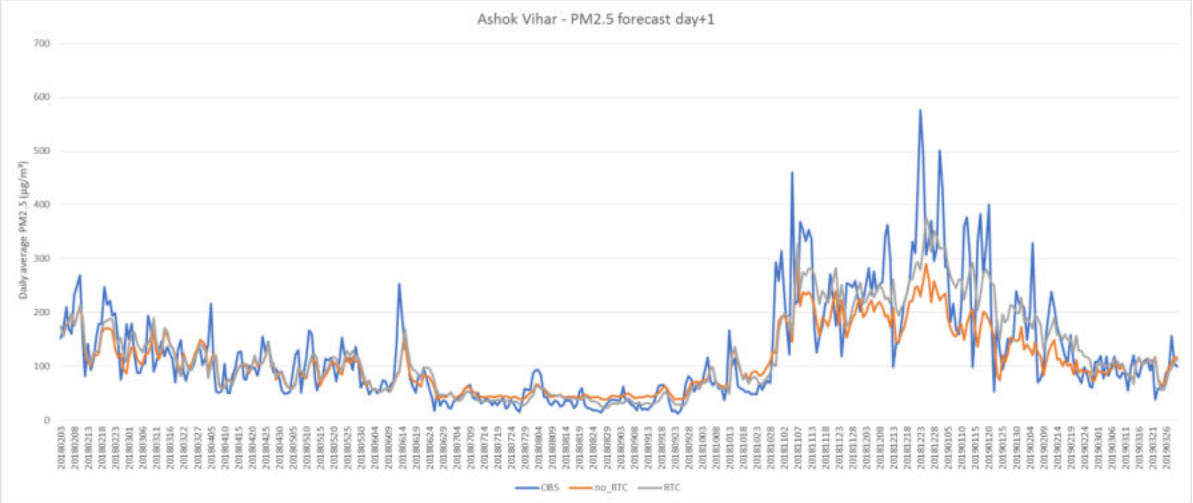


Figure 75 Comparison of daily observed PM2.5 average concentrations (yellow) at Ashok Vihar station with , predicted concentrations day+1 no RTC (blue), predicted concentrations day +1 with RTC (green).

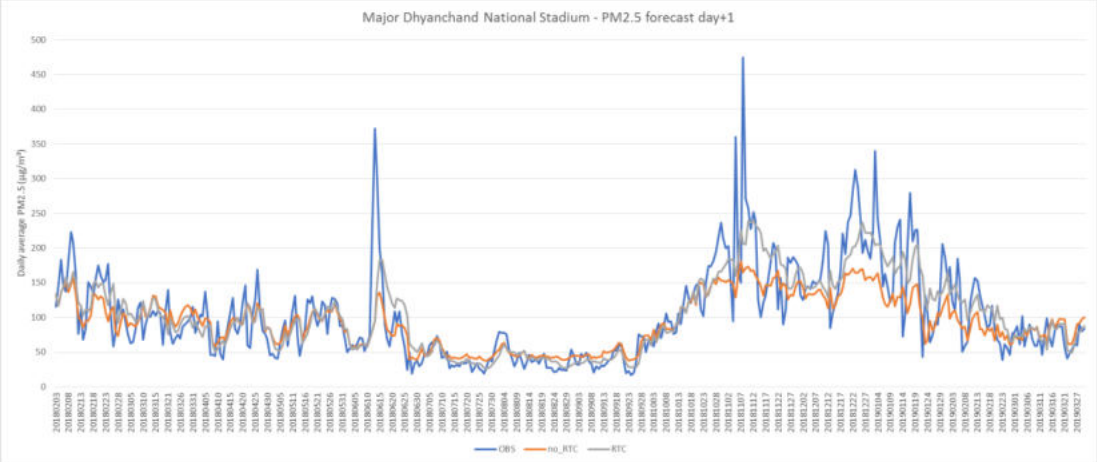


Figure 76 Comparison of daily observed PM2.5 average concentrations (yellow) at Major Dhyan Chand station with , predicted concentrations day+1 no RTC (blue), predicted concentrations day +1 with RTC (green).

Figure 76 indicates comparison of observed daily average $PM_{2.5}$ concentrations at station Major Dhyanchand with the forecasted values in different scenario: forecast concentration for day+1 with no RTC and forecast concentration for day+1 with RTC. Here also the trend was found to be similar to that of Ashok Vihar location. Forecasted concentrations with RTC option were found to be closer to the observation values throughout monitoring period. However, some of the peak values were not so well predicted by the model.

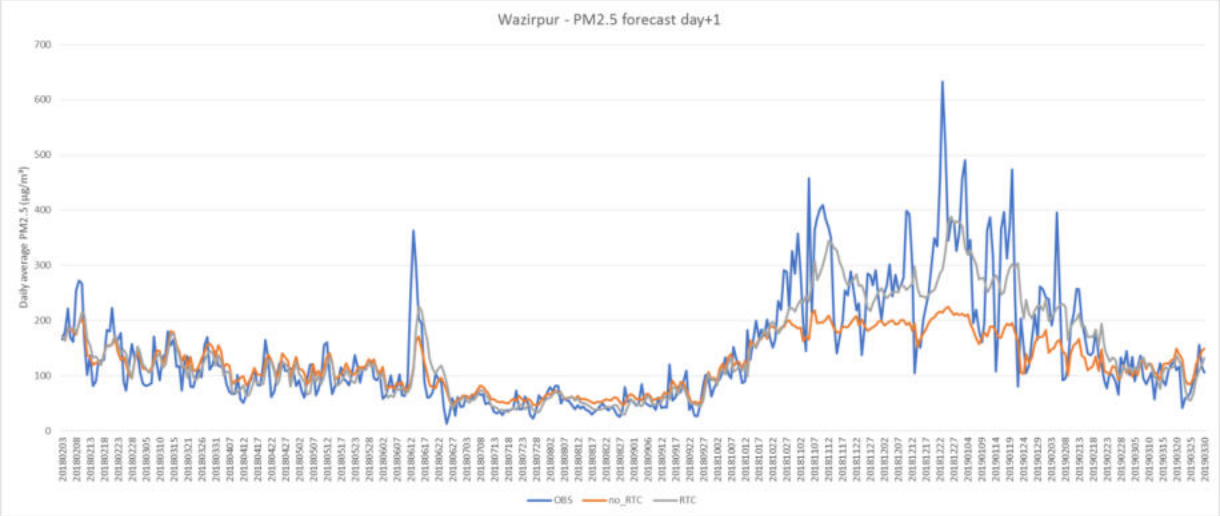


Figure 77 Comparison of daily observed PM2.5 average concentrations (yellow) at Wazirpur station with , predicted concentrations day+1 no RTC (blue), predicted concentrations day +1 with RTC (green).

Figure 77 indicates comparison of observed daily average $PM_{2.5}$ concentrations at station Wazirpur with forecasted concentrations in different forecast scenarios: forecast concentration for day+1 with no RTC and forecast concentration for day+1 with RTC. It could be interpreted that the forecasted concentration with RTC was closer to observation values throughout the monitoring period. RTC options helps in better prediction of the peaks, which are not so well predicted in the model.

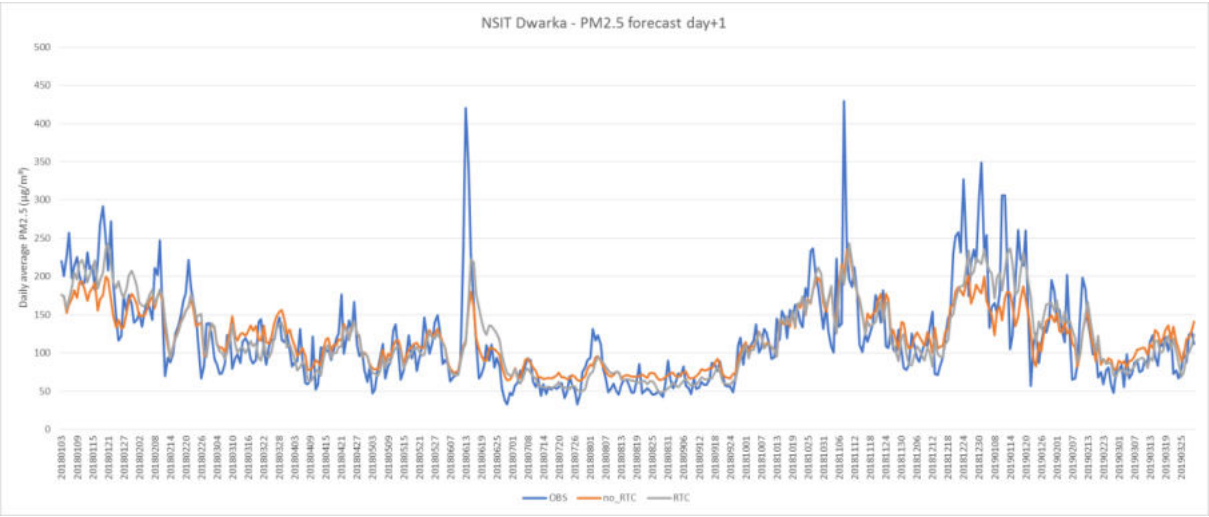


Figure 78 Comparison of daily observed PM2.5 average concentrations (yellow) at NSIT Dwarka station with , predicted concentrations day+1 no RTC (blue), predicted concentrations day +1 with RTC (green).

Figure 78 represents the daily average $PM_{2.5}$ concentrations at station NSIT Dwarka considering observed concentration and different forecast scenario: forecast concentration for day+1 with no RTC and forecast concentration for day+1 with RTC. The trend shown in this location is also similar to that observed for other locations in which forecasted

concentration with RTC was close to observation value throughout monitoring period, except for very high pollution concentrations. Both the scenarios are showing good correlation with the observed concentrations for all months except for a few very high observations RTC helps improving the forecasts

5.6 Conclusion

- Correlation analysis showed strong correlation of pollutant concentrations with concentrations of previous days and meteorological variables like boundary layer height, temperature, precipitation etc and hence integrated into the forecasting model for predictive purpose.
- Forecasting model satisfactorily predicts pollutant concentrations at most stations.
- RTC -Real time corrections have been tested and found to to improve the model performance .
- Correlation with the number of fire events with the pollutant concentrations was found to be is fairly limited when full year dataset has been considerd. Therefore adding the fire events (average fire events over the past 3 days) as extra variables does not make a difference for the validation statistics.
- Longer time series to train the model, as well as improved meteorological datasets can improve the model configuration and may lead to better prediction efficiencies.
- Reducing the variables (concentration_day0 until 9am, boundary layer height, total precipitation, fire events) is not adding any benefit to the outputs, in fact it deteriorates performance.

6. Training Program on spatial mapping and forecasting models

A one day training program was conducted for officials of CPCB by technical experts from VITO, Belgium in March 2019 at CPCB office in Delhi to understand the basic concepts and use of spatial (RIO) and forecasting (OVL) models. The training program was attended by officials from CPCB and TERI. The basic objective of the training program was to build the capacity within CPCB in setting up spatial and forecasting models.

Dr Shukla from CPCB welcomed the experts from VITO and explained the need for conducting such training program in India.

Dr Sumit Sharma from TERI explained the project objectives, TERI's association with VITO and the models developed by VITO for spatial mapping and forecasting of air pollutants.

Ms Byth Lisa from VITO explained at length their experience in air quality modelling in the context of the EU AQ directive: application of models for forecasting, assessment and planning.

This was followed by a brief presentation Dr Stijn Janssen from VITO on the OVL and RIO models. The specific goal of his presentation was to make the attendees understand (in basic terms) how the models work and what data is required to run the model.

This was followed by presentation by Mr Stijn Vranckx, VITO on the configuration of the models for Delhi. The main objectives of his presentation was to understand the participants about the data requirements to set-up the model, the steps taken to set-up the model and how the data impacts the performance. He explained the participants starting from the data clean-up to the final configuration and the validation done to-date. He also presented on how to use the user interface for the manual daily check of the performance of OVL.

7. Conclusion

Spatial mapping and forecasting of air pollutants is very important to understand and plan mitigative actions for control. They can also be useful in drafting adaptive strategies to reduce the exposure of pollutants to larger populations. In this study, TERI with support from VITO customised a spatial mapping and a forecasting model for the city of Delhi. These models have been previously used in several parts of the world. During the course of the study, both spatial mapping and forecasting model has been successfully setup, validated, and tested for the city of Delhi. Along with the model setup, a training program was also organised for the CPCB officials to train the officials for regular use of the models.

In case of the spatial model, PM has been more adequately mapped than other gaseous pollutants and further research is required to develop parameters for adequate spatial interpolation of these pollutants. The spatial maps prepared using the validated RIO model shows distinct seasonal variation, and also highlights the hotspot regions in the city. The whole city is under the influence of high pollutant levels depicting high background concentrations and widespread local sources (like vehicles). However, certain industrial regions are found to be even more polluted and showed up as hotspots. Other than industries, the whole western and north western of Delhi, North of Ashok Vihar, Rohini, Timarpur, area around stations of Janhangir Puri and Wazirpur were also identified as the hotspots in the city.

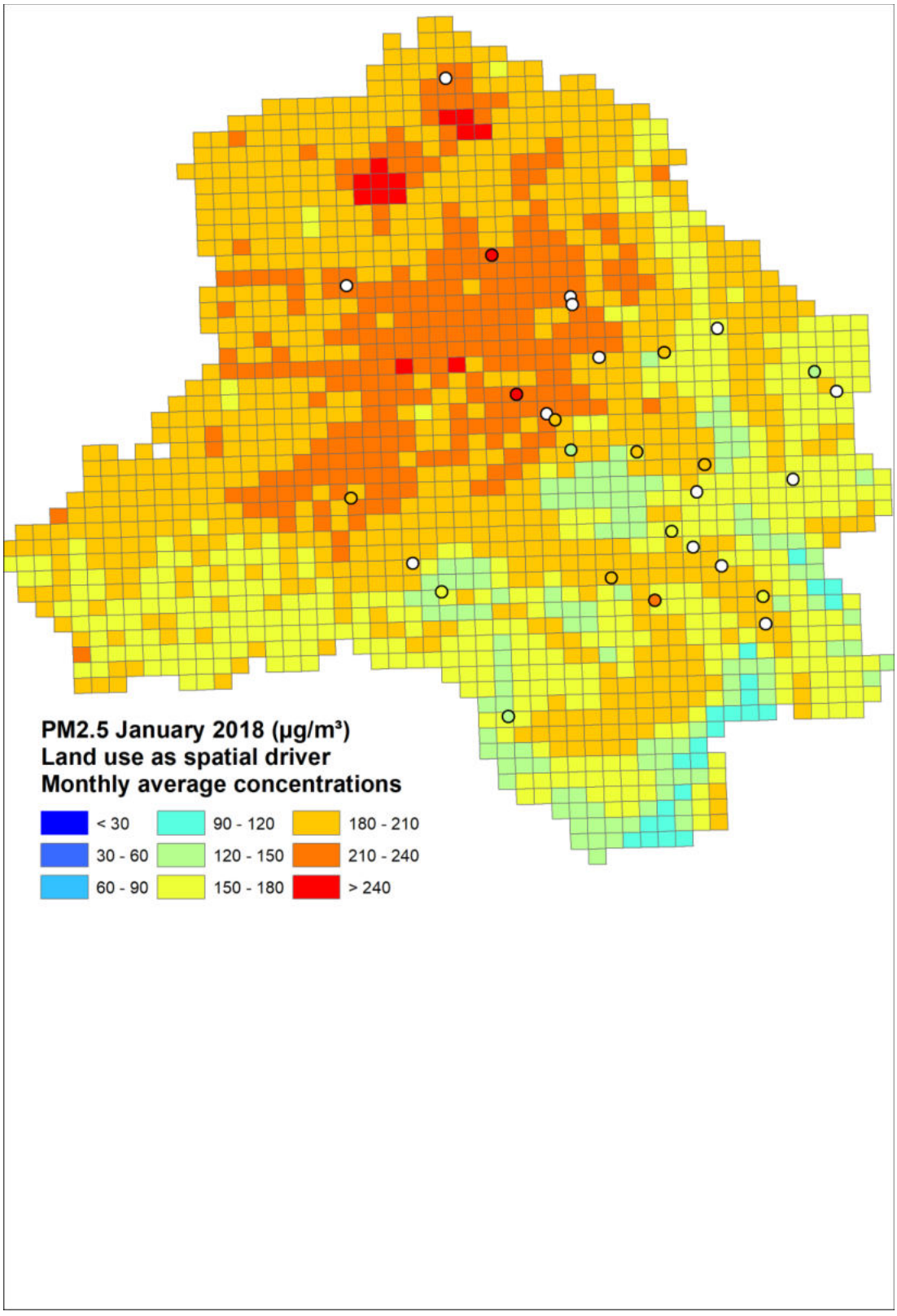
Forecasting of pollutants using the OVL model shows satisfactory performance for most part of the year, except for some peaks of pollutants concentrations which could not be so well reproduced by the model. The performance of the model for prediction of peaks improves considerably with the usage of RTC option. Use of longer term datasets (about 3-4 years) could further improve the model performance for the city.

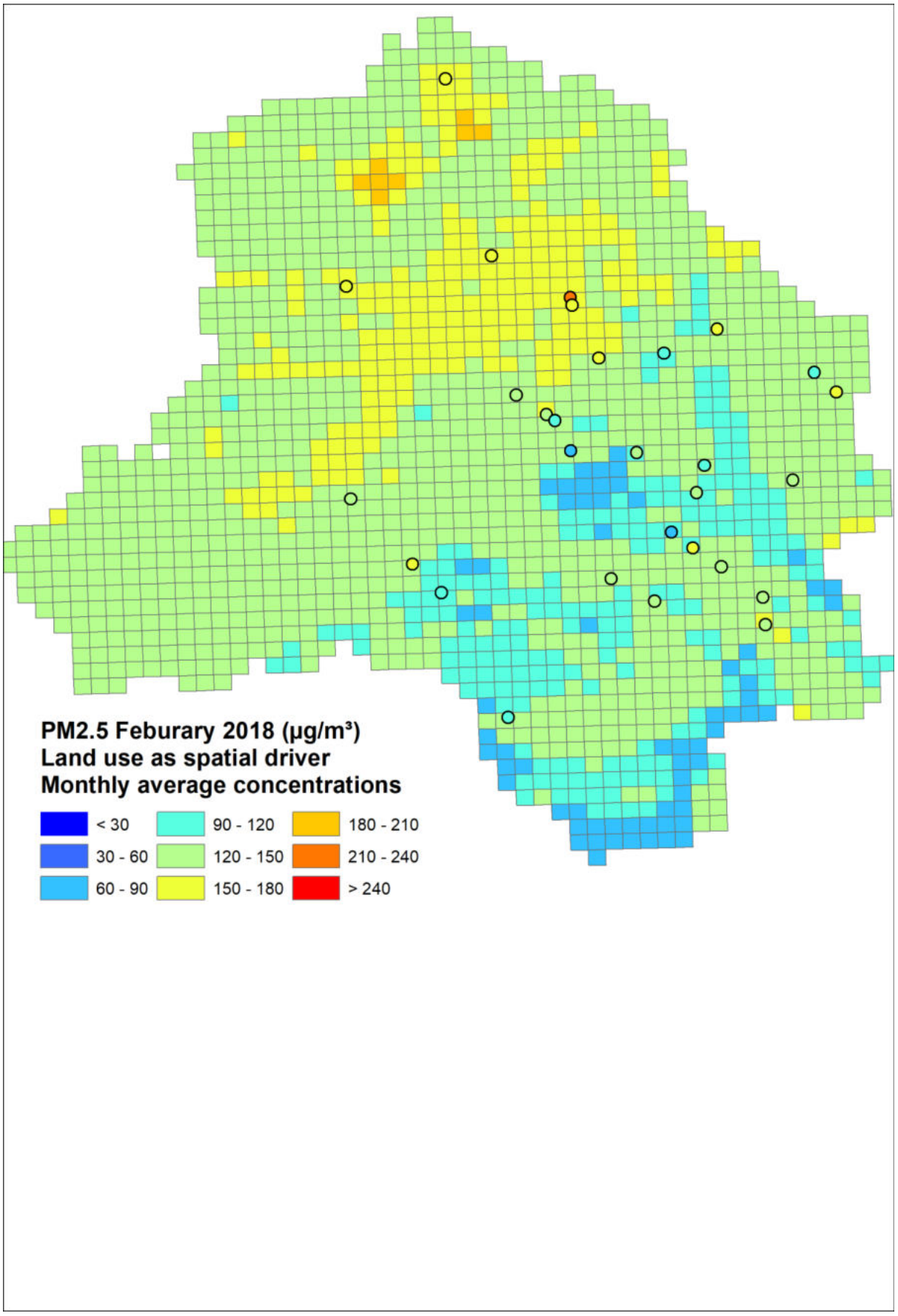
The models are satisfactorily tested and ready to be used for demonstrative purposes for the city of Delhi. Moreover, the models can now be replicated in other cities after required setting up, validation, and testing procedures.

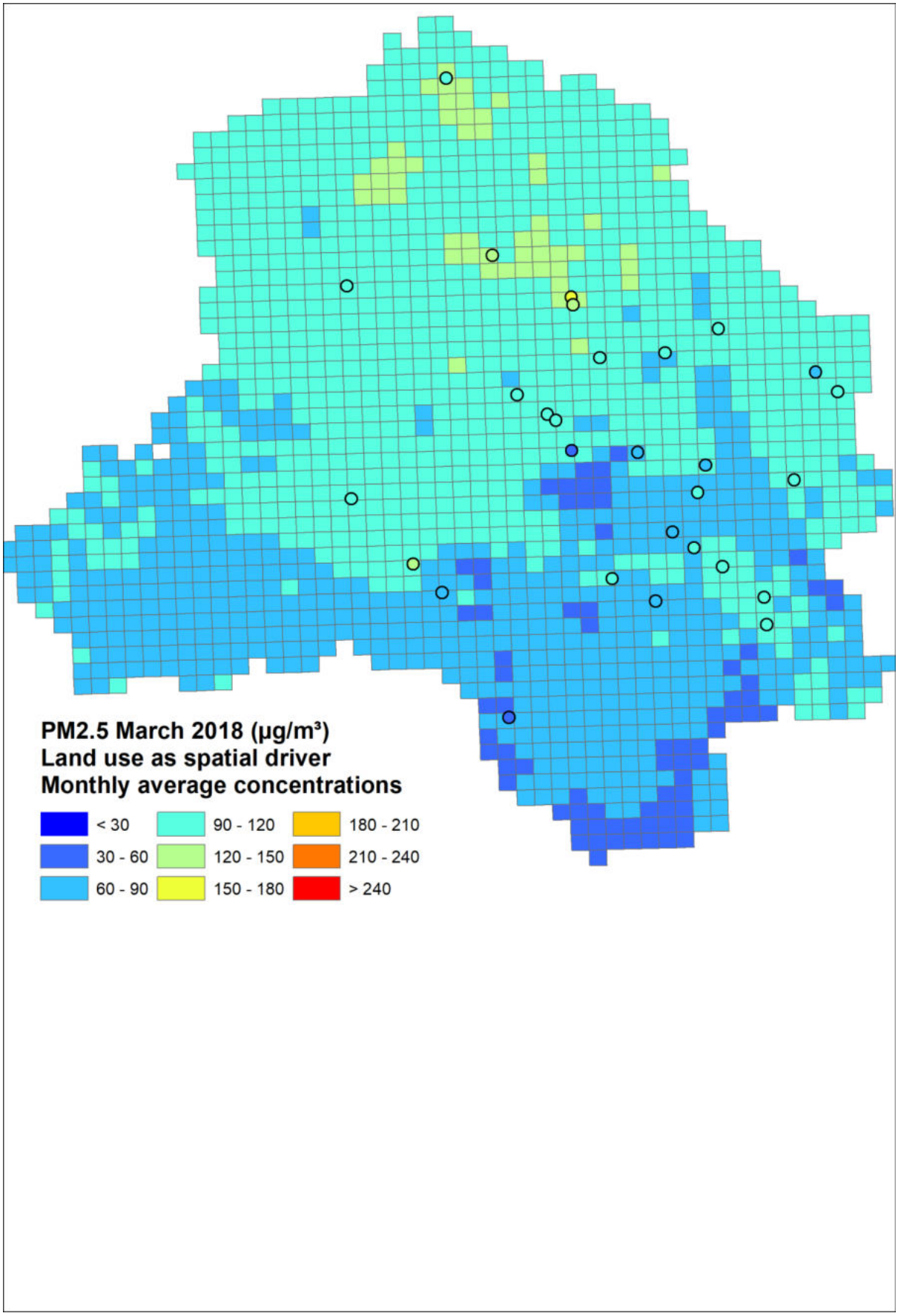
8. References

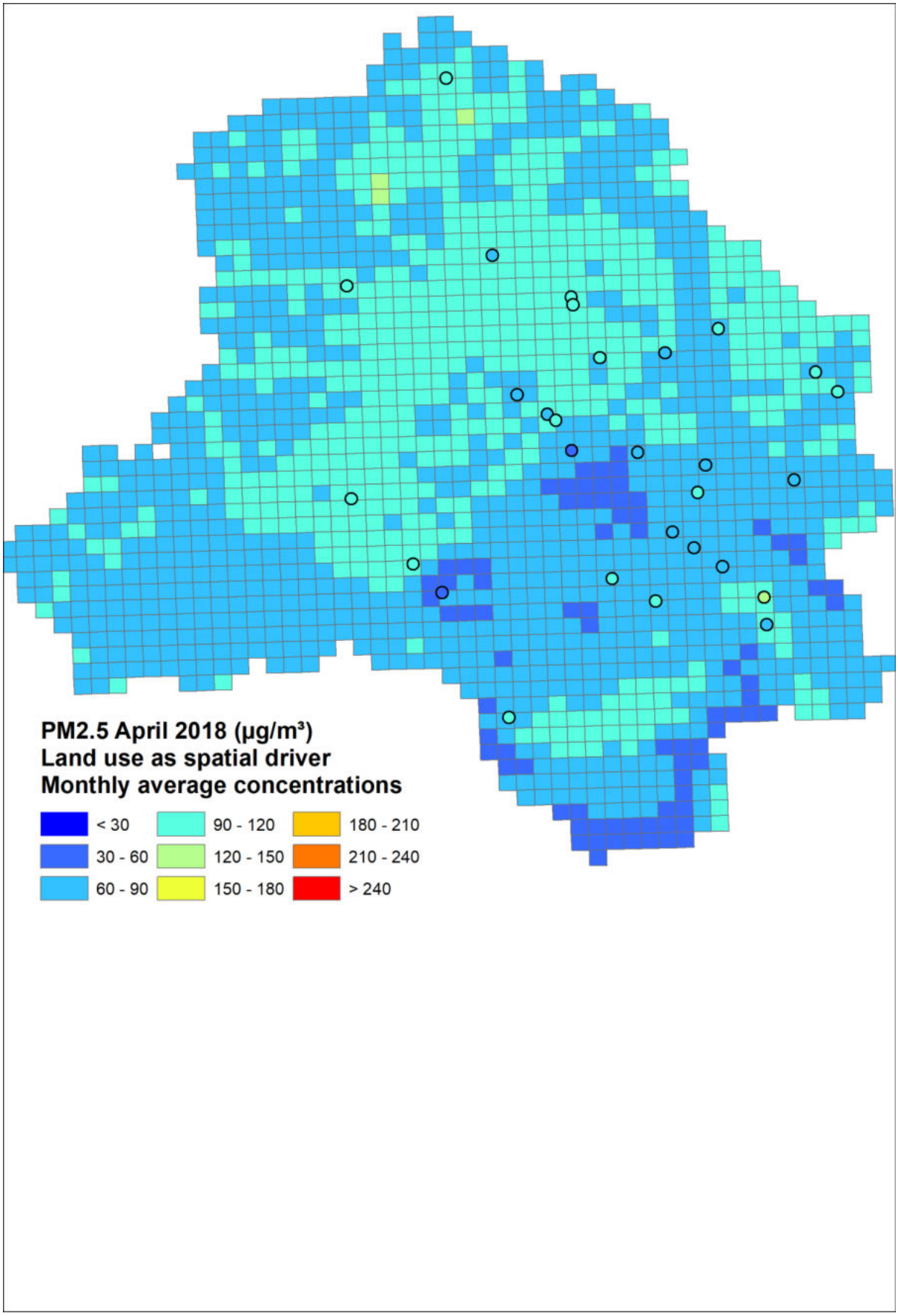
1. Cropper, M.L., Simon, N.B., Alberini, A., Arora, S., Sharma, P., 1997. The health benefits of air pollution control in Delhi. Am. J. Agric. Econ. 79, 1625 - 1629.
2. Dholakia, H.H., Purohit, P., Rao, S., Garg, A., 2013.Impact of current policies on future air quality and health outcomes in Delhi, India. Atmos. Environ. 75, 241- 248.
3. Guttikunda, S.K., Goel, R., 2013. Health impacts of particulate pollution in a megacity Delhi
4. HEI, 2018, Burden of Disease Attributable to Major Air Pollution Sources in India, January 2018.
5. https://www.healtheffects.org/system/files/GBD-MAPS-SpecRep21-India-revised_0.pdf. Accessed on 30 Jan 2018.
6. http://www.cpcb.gov.in/CAAQM/mapPage/frmindiamap.aspxcpcb.nic.in
7. IHME, 2013. The Global Burden of Disease 2010: Generating Evidence and Guiding Policy. Institute for Health Metrics and Evaluation, Seattle, USA. India. Environ. Dev. 6, 8 - 20.
8. Kandlikar, M., Ramachandran, G., 2000. The causes and consequences of particulate air pollution in urban India: a synthesis of the science. Annu. Rev. Energy Environ. 25, 629 - 684.
9. Cropper, M.L., Simon, N.B., Alberini, A., Arora, S., Sharma, P., 1997. The health benefits of air pollution control in Delhi. Am. J. Agric. Econ. 79, 1625 - 1629.
10. Dholakia, H.H., Purohit, P., Rao, S., Garg, A., 2013.Impact of current policies on future air quality and health outcomes in Delhi, India. Atmos. Environ. 75, 241- 248.
11. Guttikunda, S.K., Goel, R., 2013. Health impacts of particulate pollution in a megacity Delhi
12. HEI, 2018, Burden of Disease Attributable to Major Air Pollution Sources in India, January 2018.
13. https://www.healtheffects.org/system/files/GBD-MAPS-SpecRep21-India-revised_0.pdf. Accessed on 30 Jan 2018.
14. http://www.cpcb.gov.in/CAAQM/mapPage/frmindiamap.aspxcpcb.nic.in
15. IHME, 2013. The Global Burden of Disease 2010: Generating Evidence and Guiding Policy. Institute for Health Metrics and Evaluation, Seattle, USA. India. Environ. Dev. 6, 8 - 20.
16. Kandlikar, M., Ramachandran, G., 2000. The causes and consequences of particulate air pollution in urban India: a synthesis of the science. Annu. Rev. Energy Environ. 25, 629 - 684.
17. B Roberto et.al. A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems, monthly weather review, volume 133, 2004

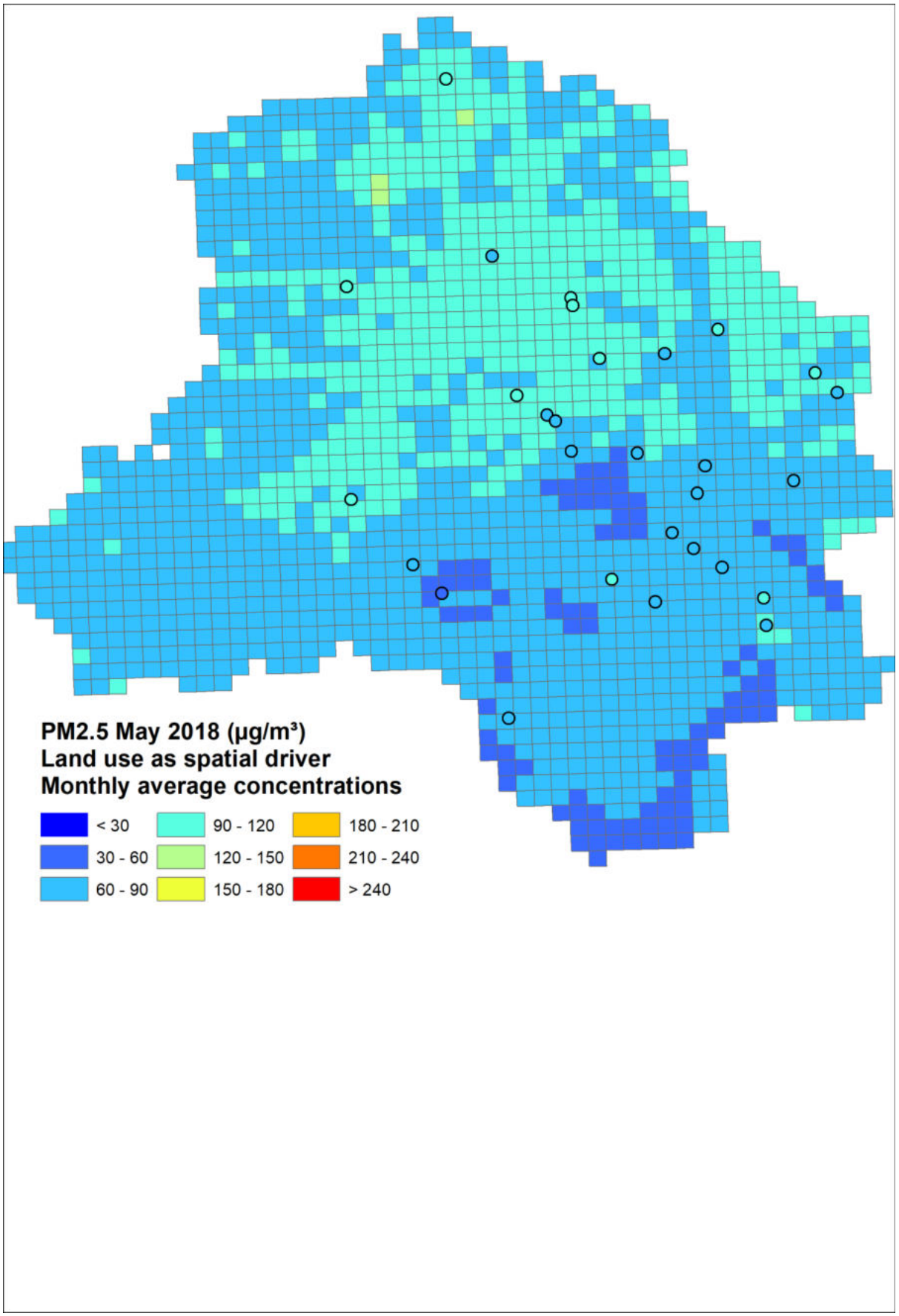
Annexure I: Monthly spatial maps for $PM_{2.5}$ from January 2018 to March 2019

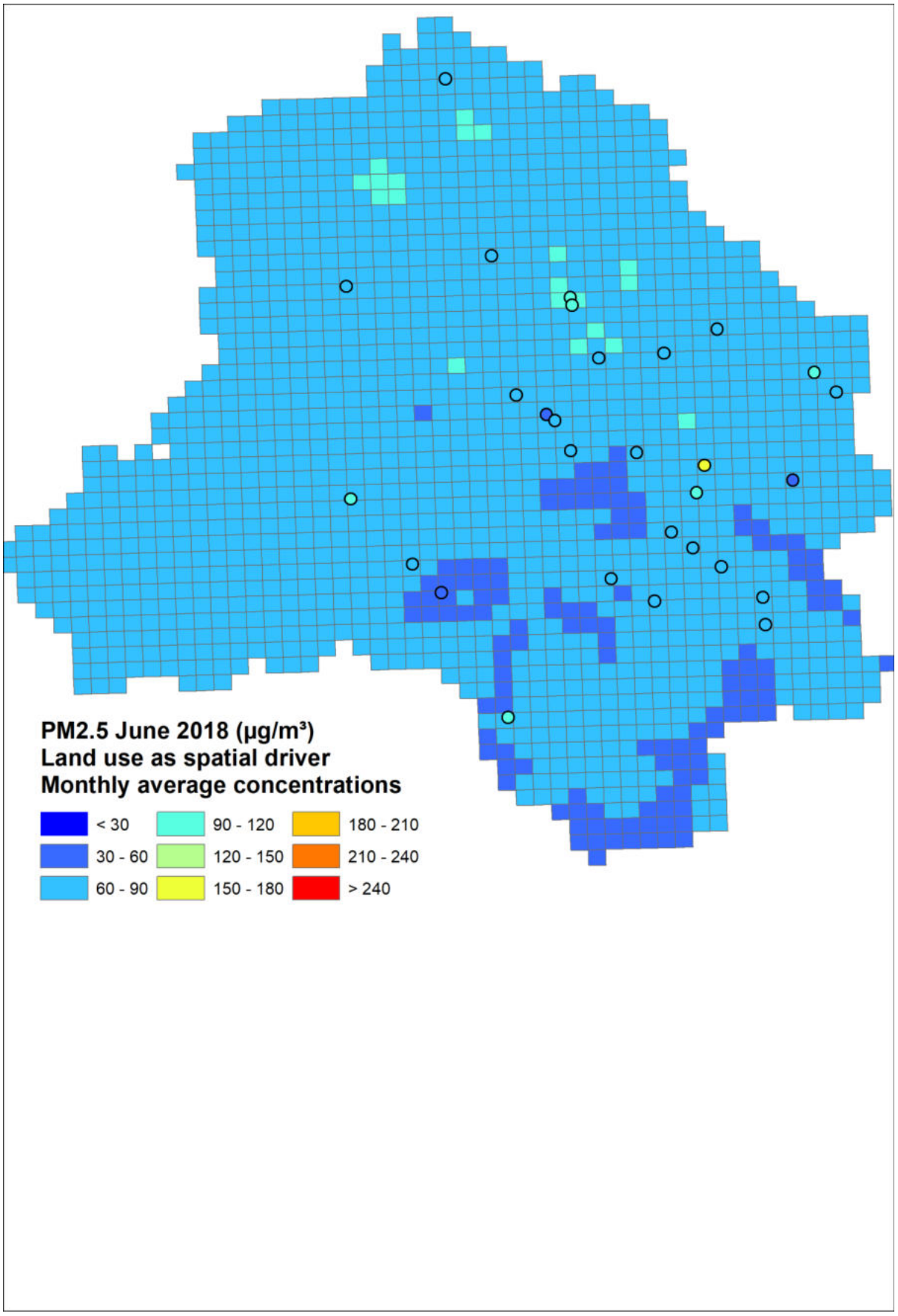


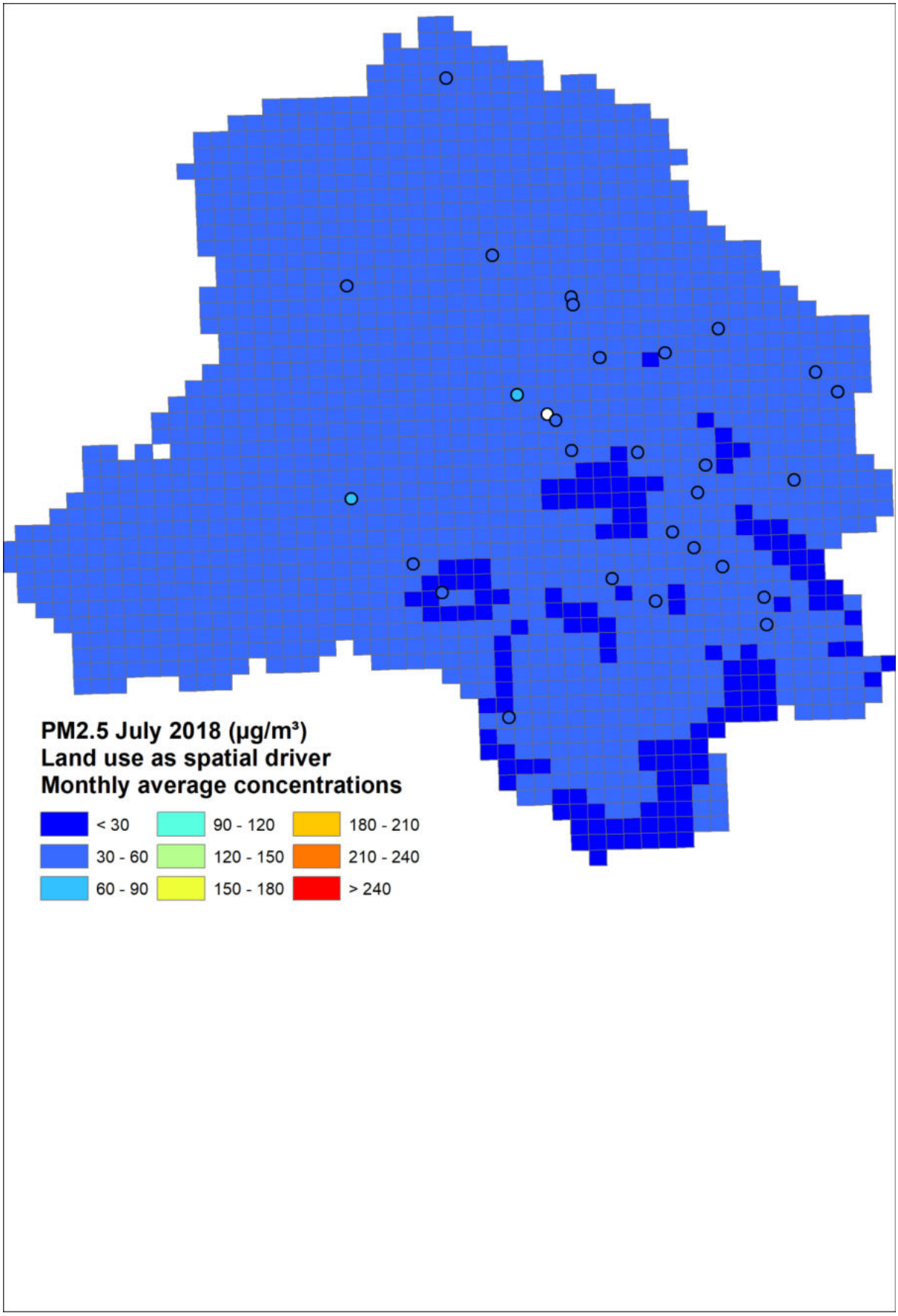

PM2.5 Feburary 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

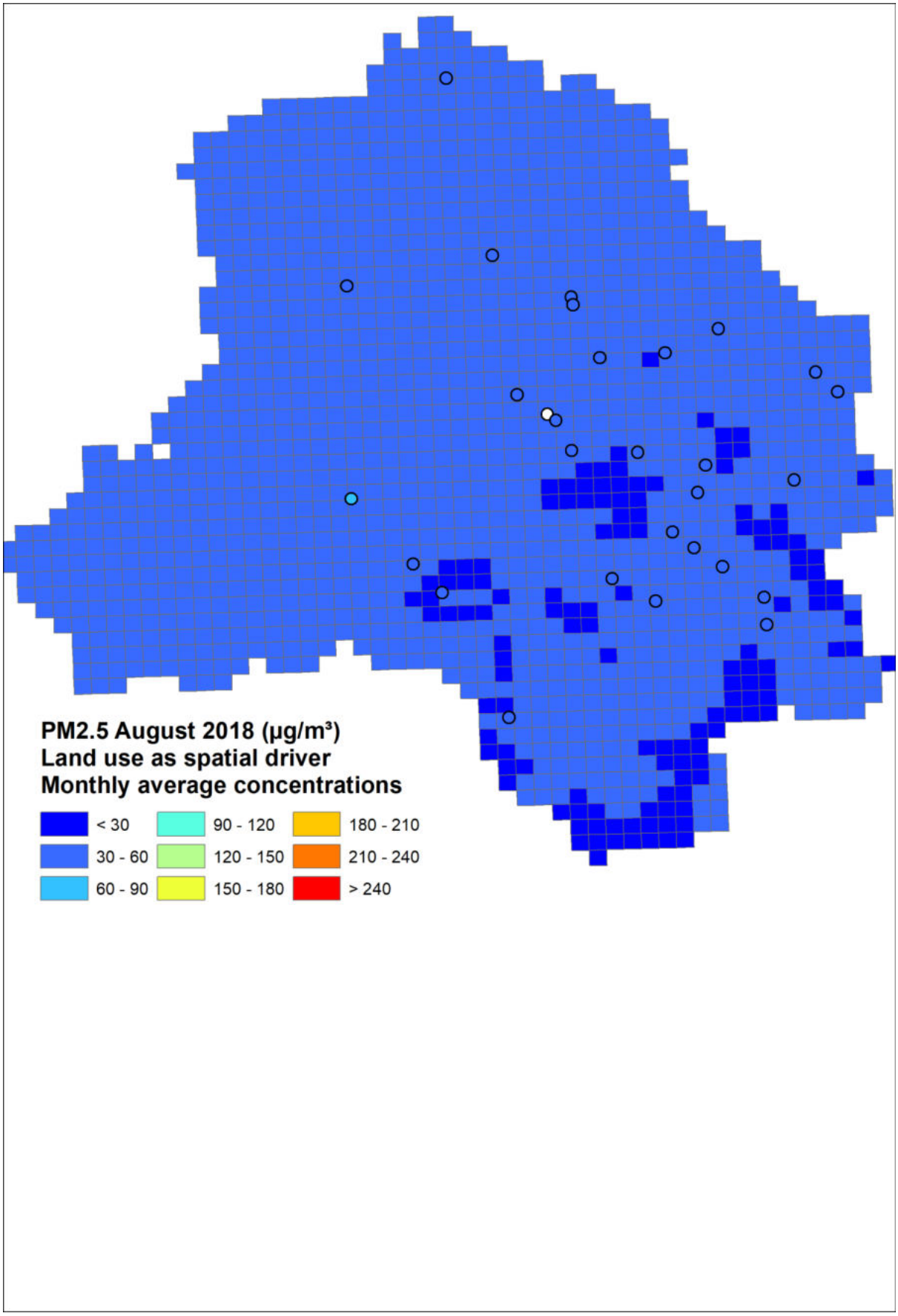

PM2.5 March 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

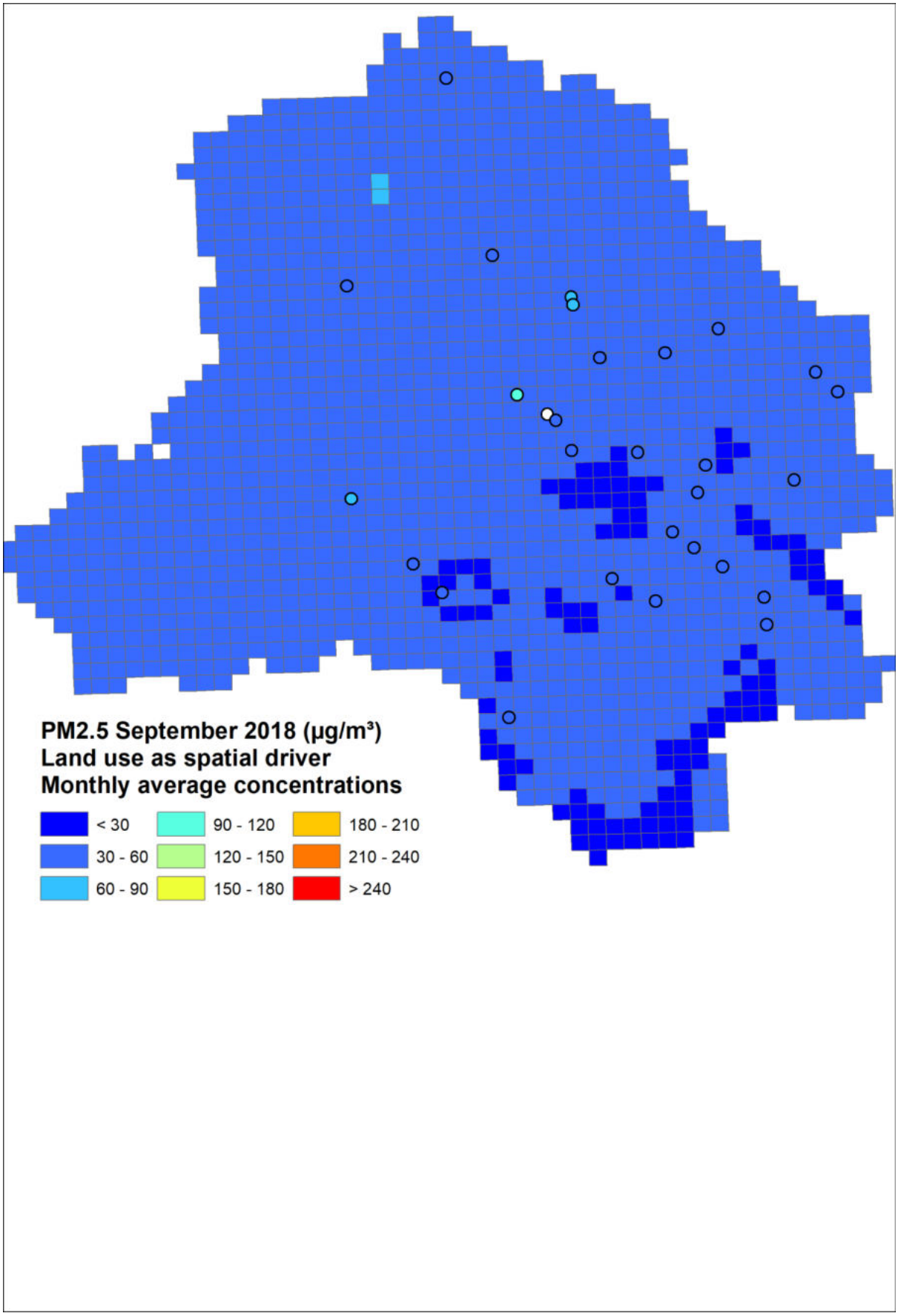

PM2.5 April 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

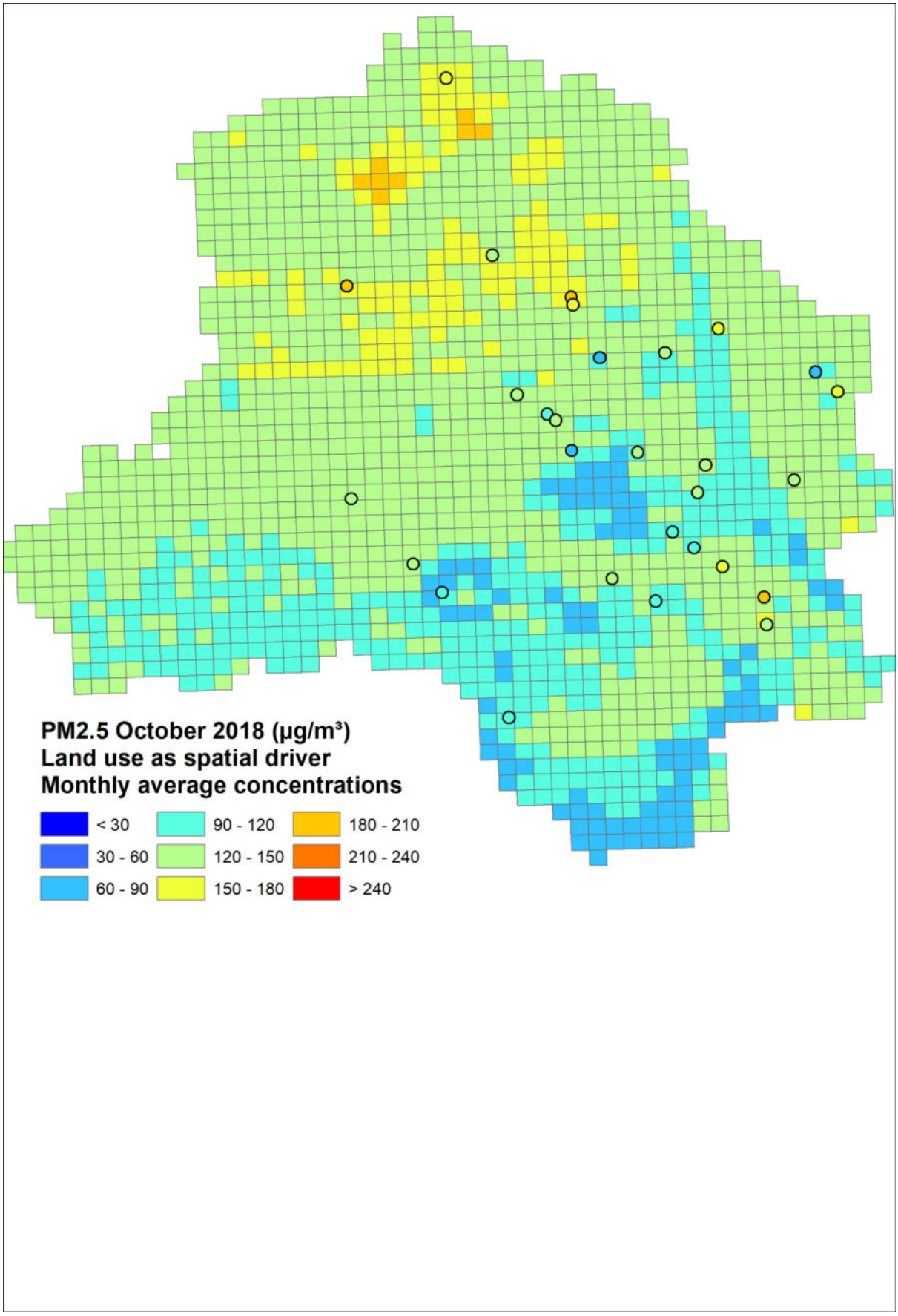

PM2.5 May 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

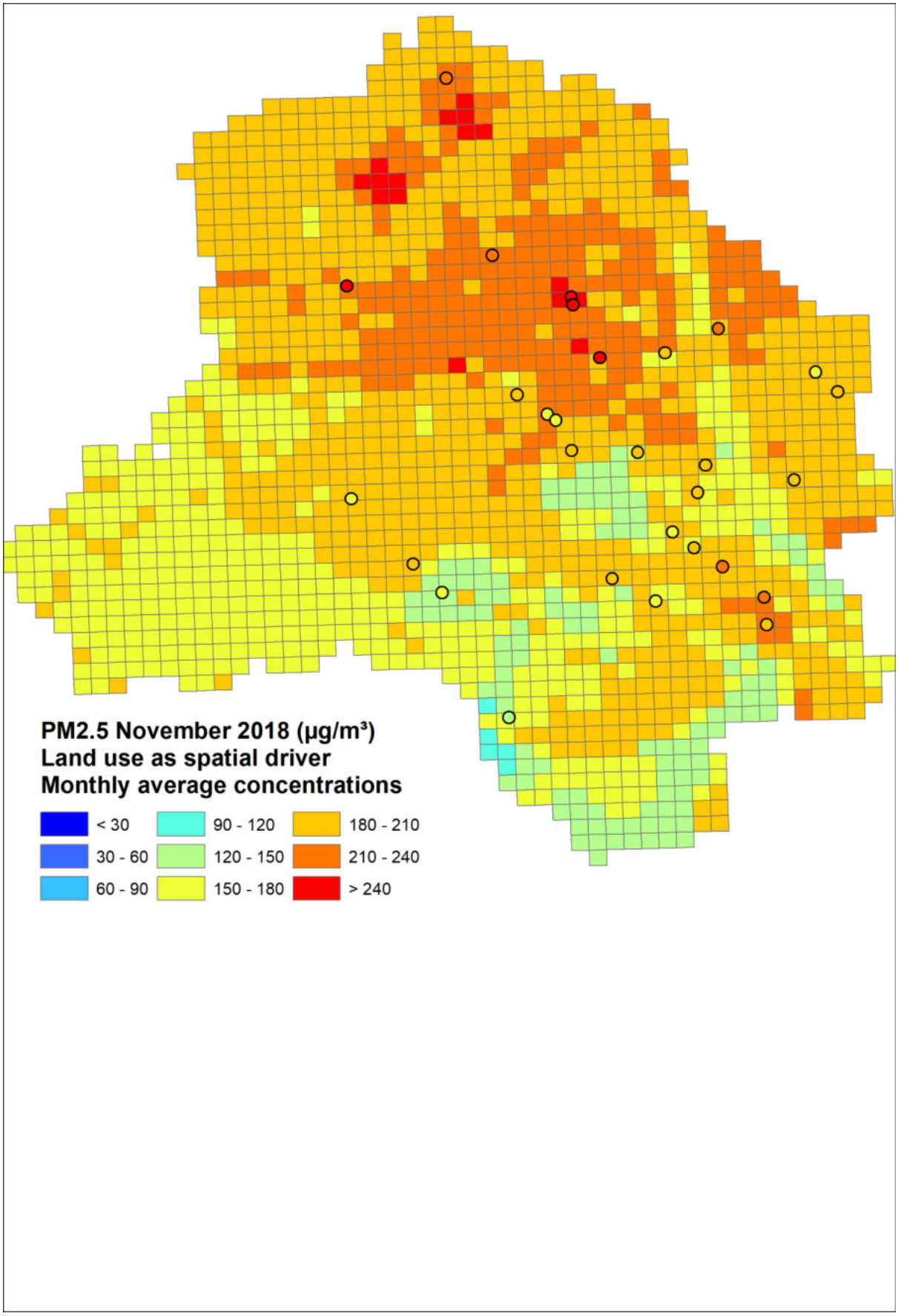

PM2.5 June 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

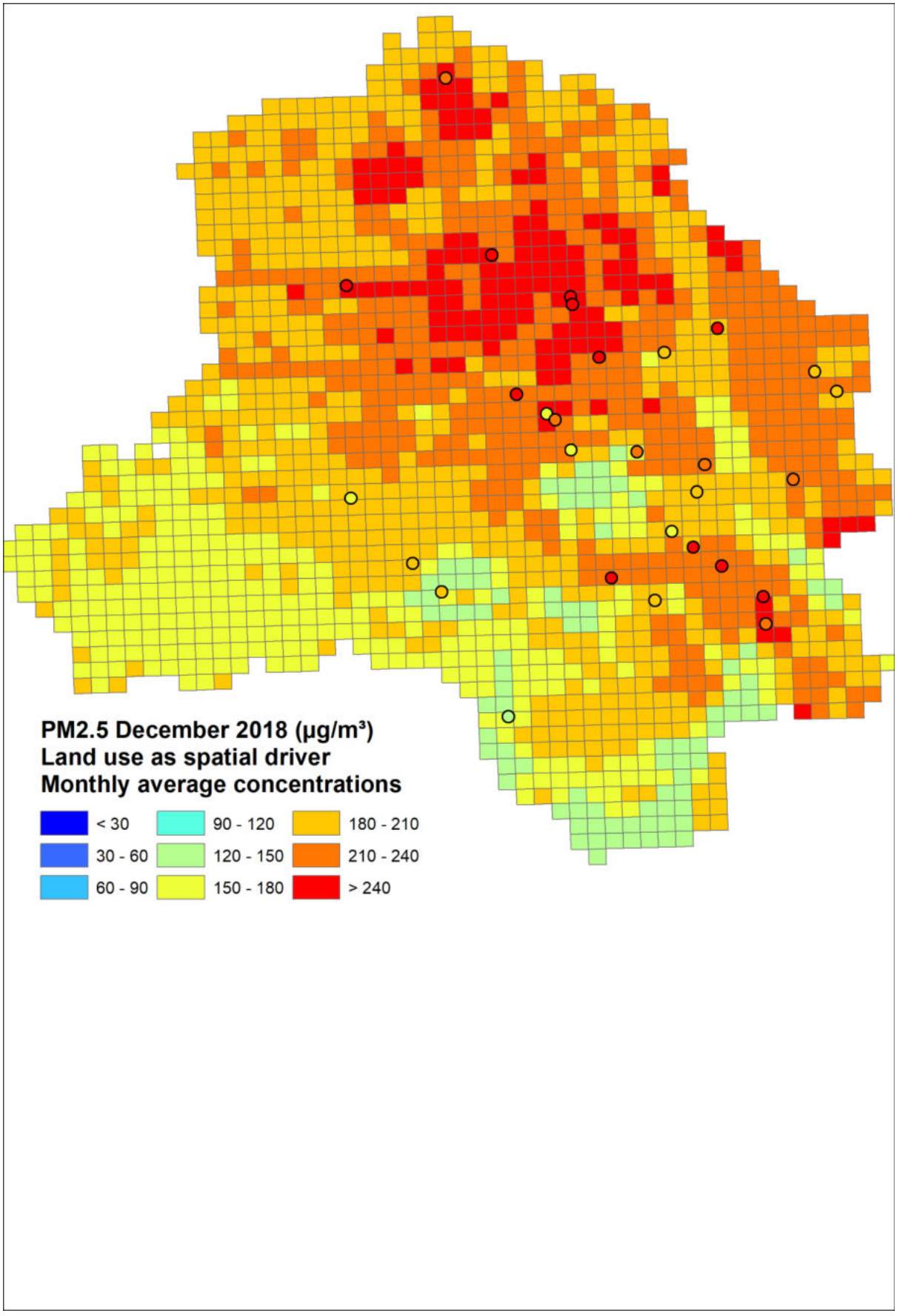

PM2.5 July 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

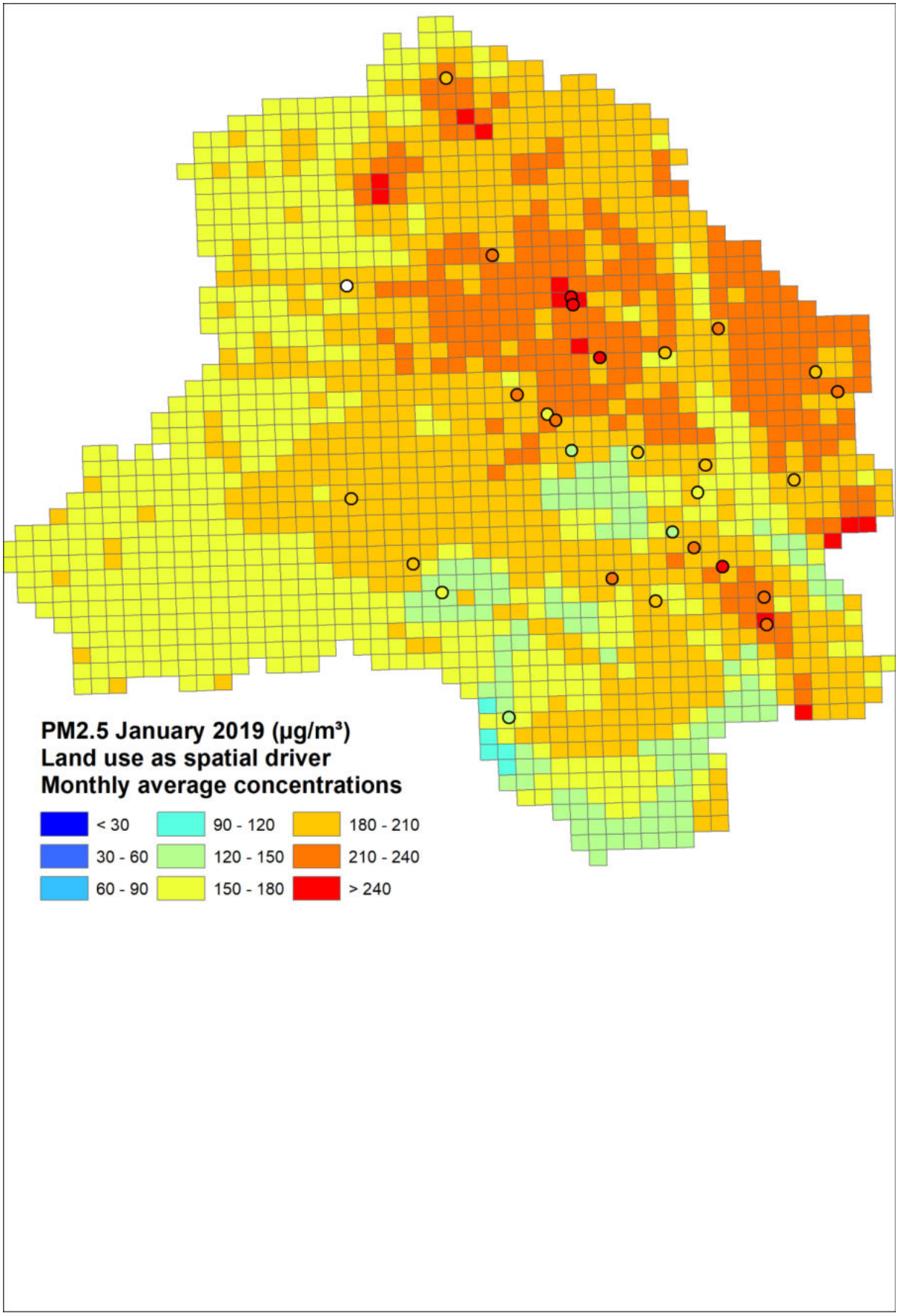

PM2.5 August 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

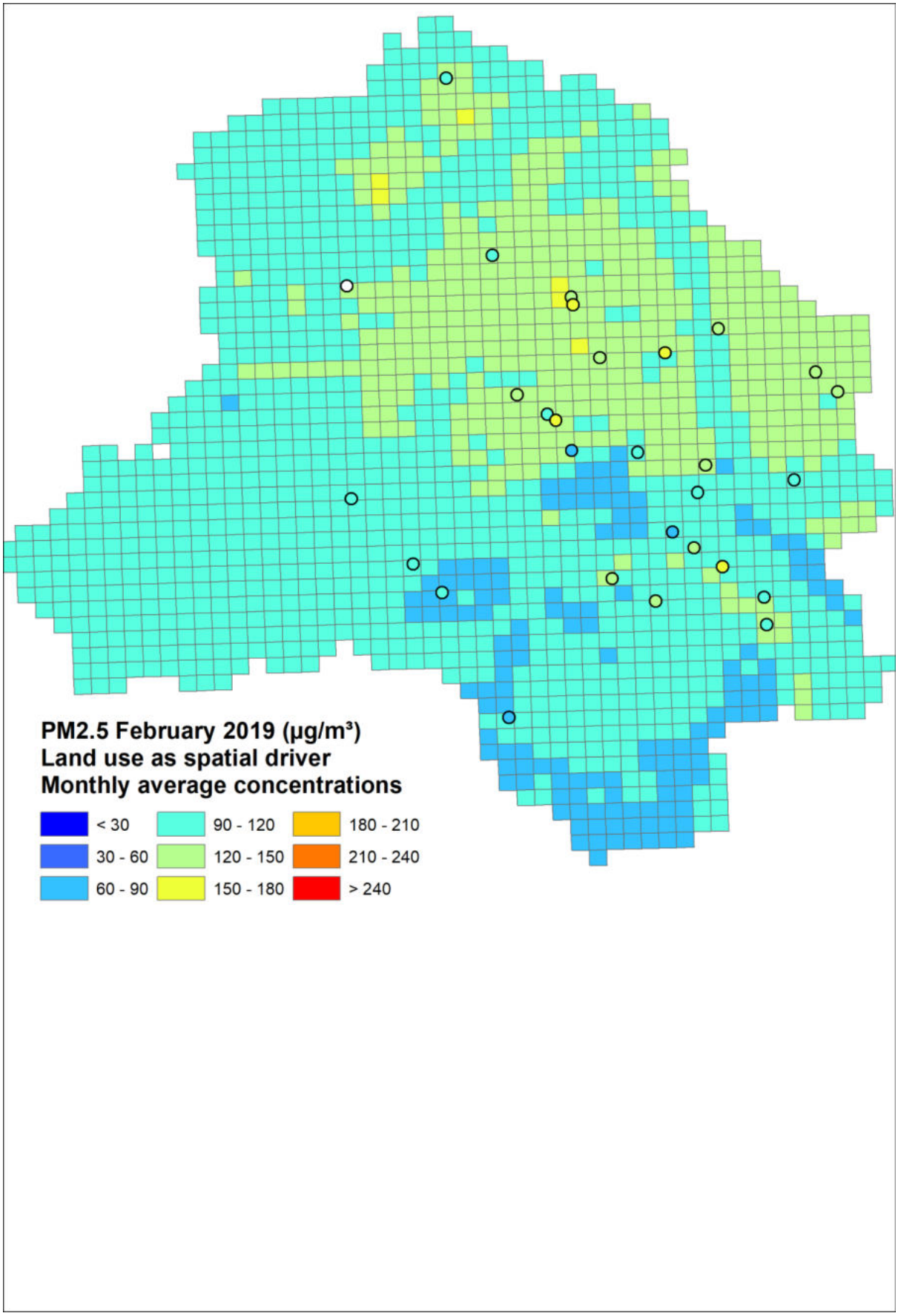

PM2.5 September 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

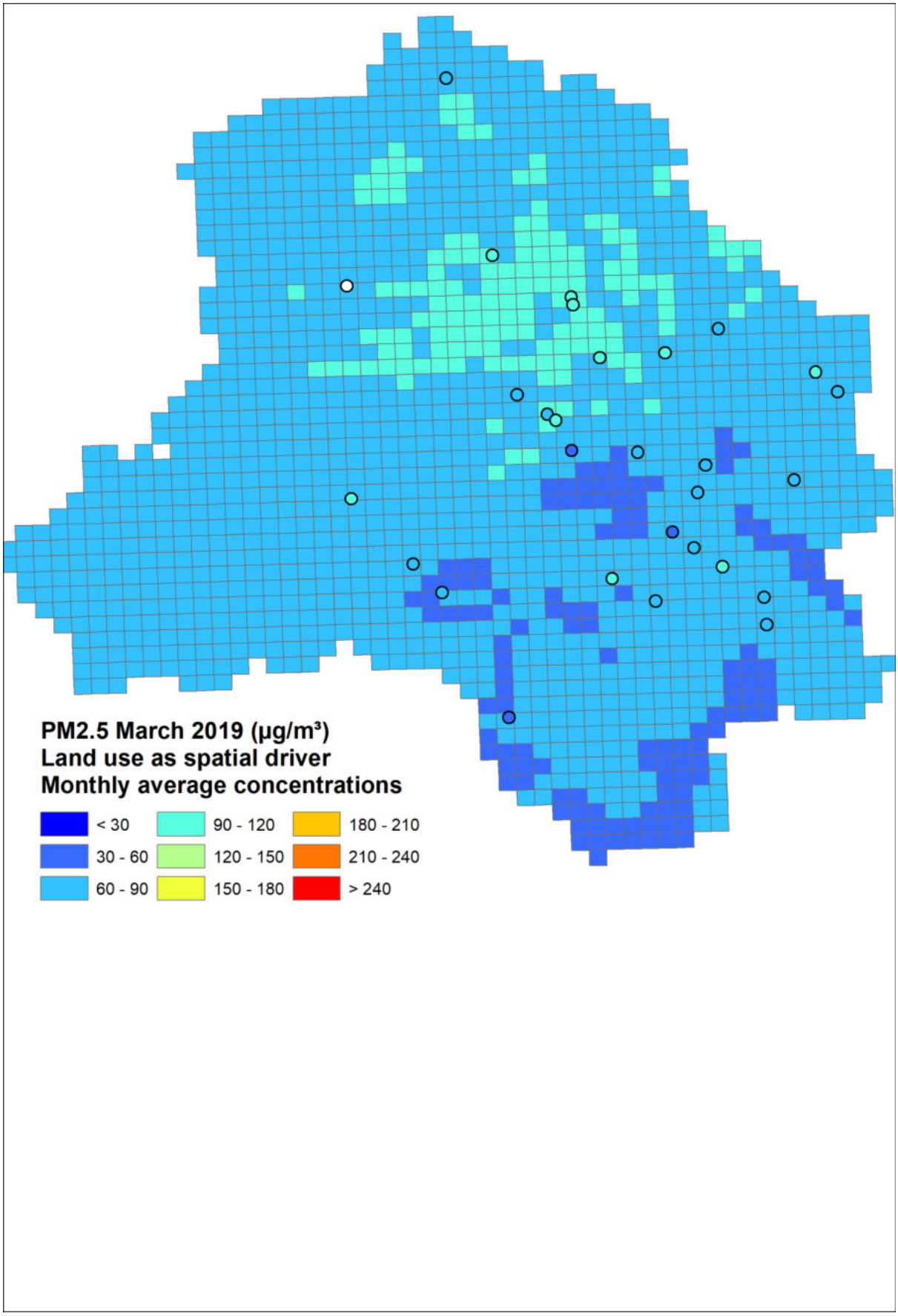

PM2.5 October 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

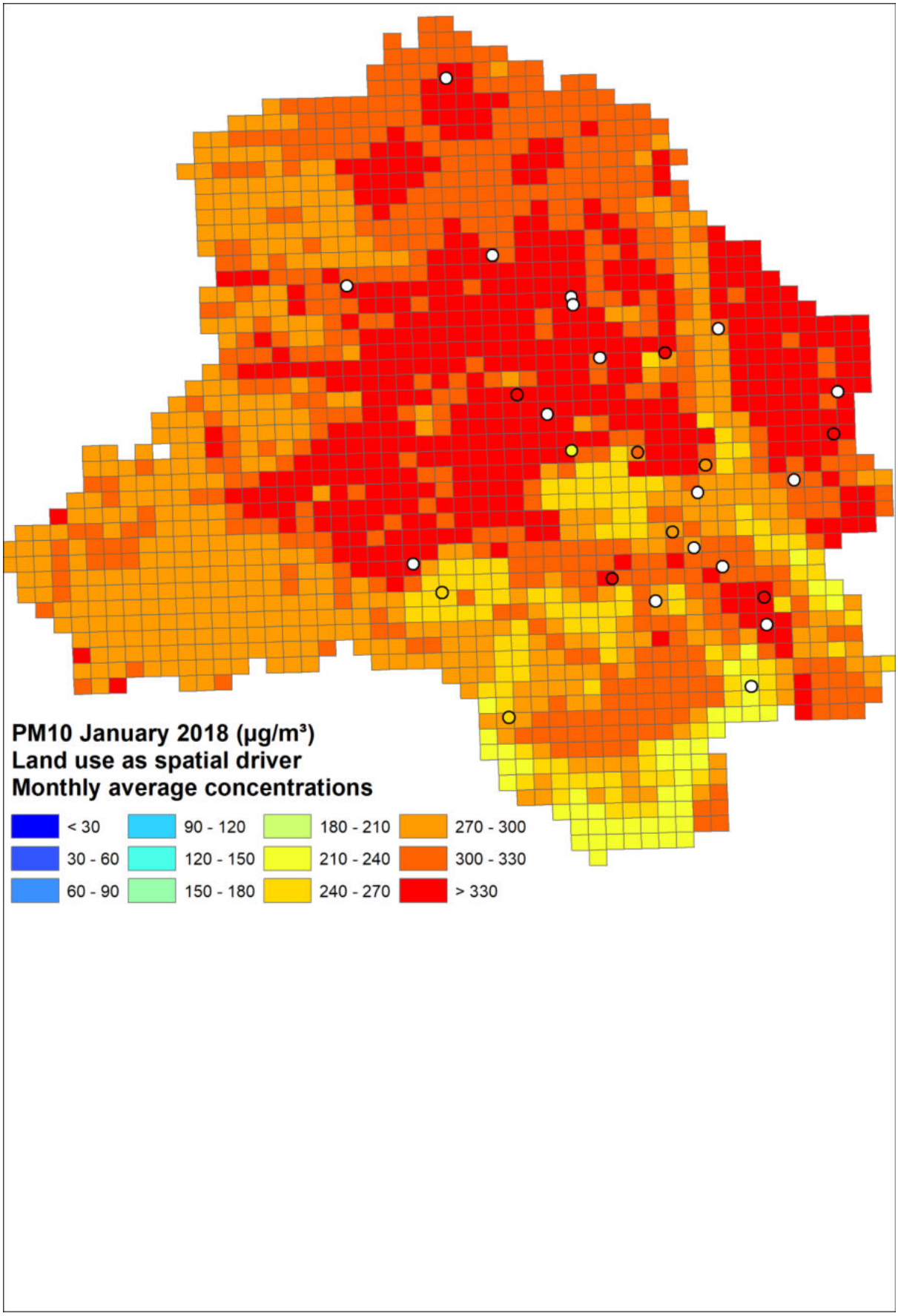

PM2.5 November 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

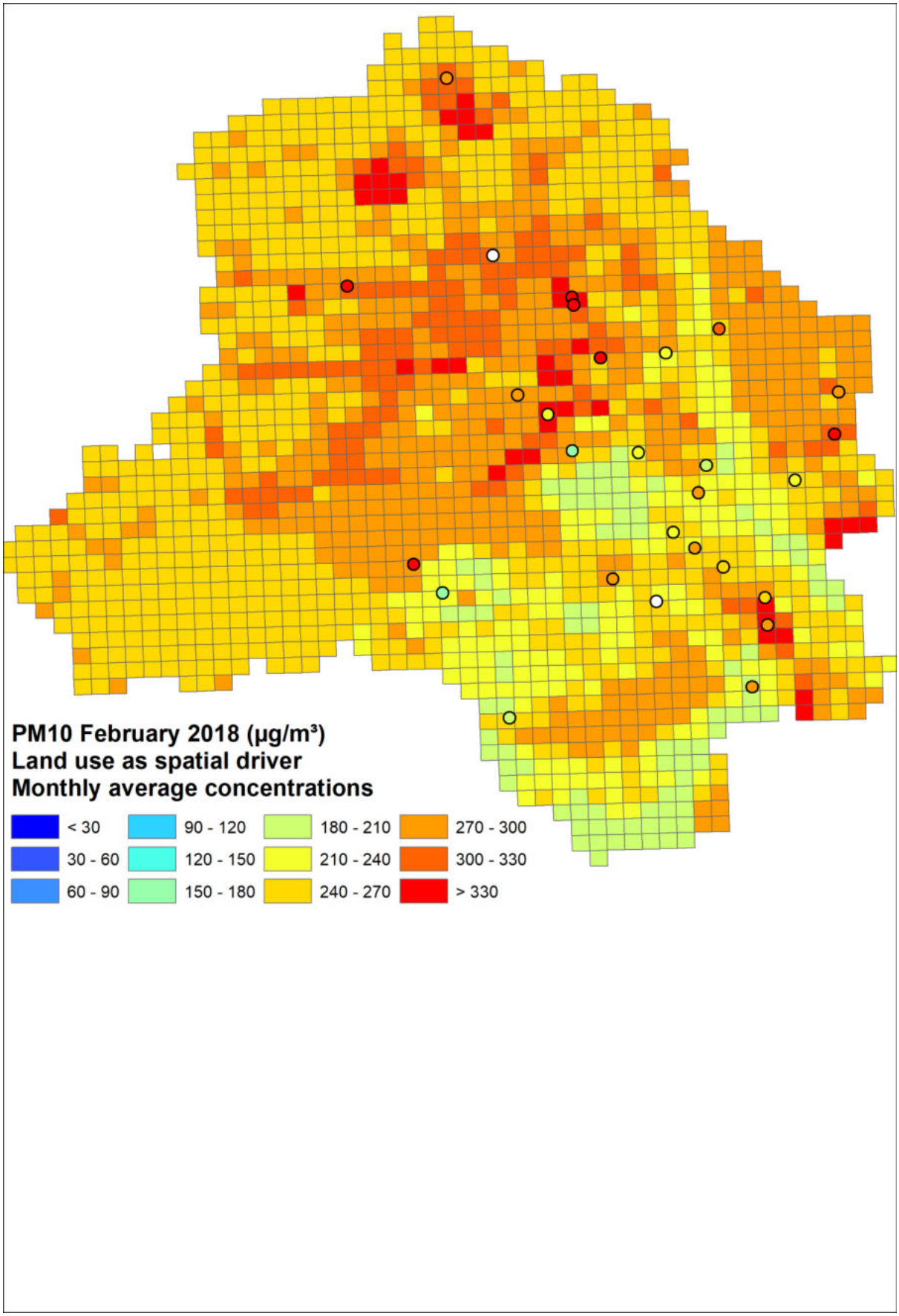

PM2.5 December 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

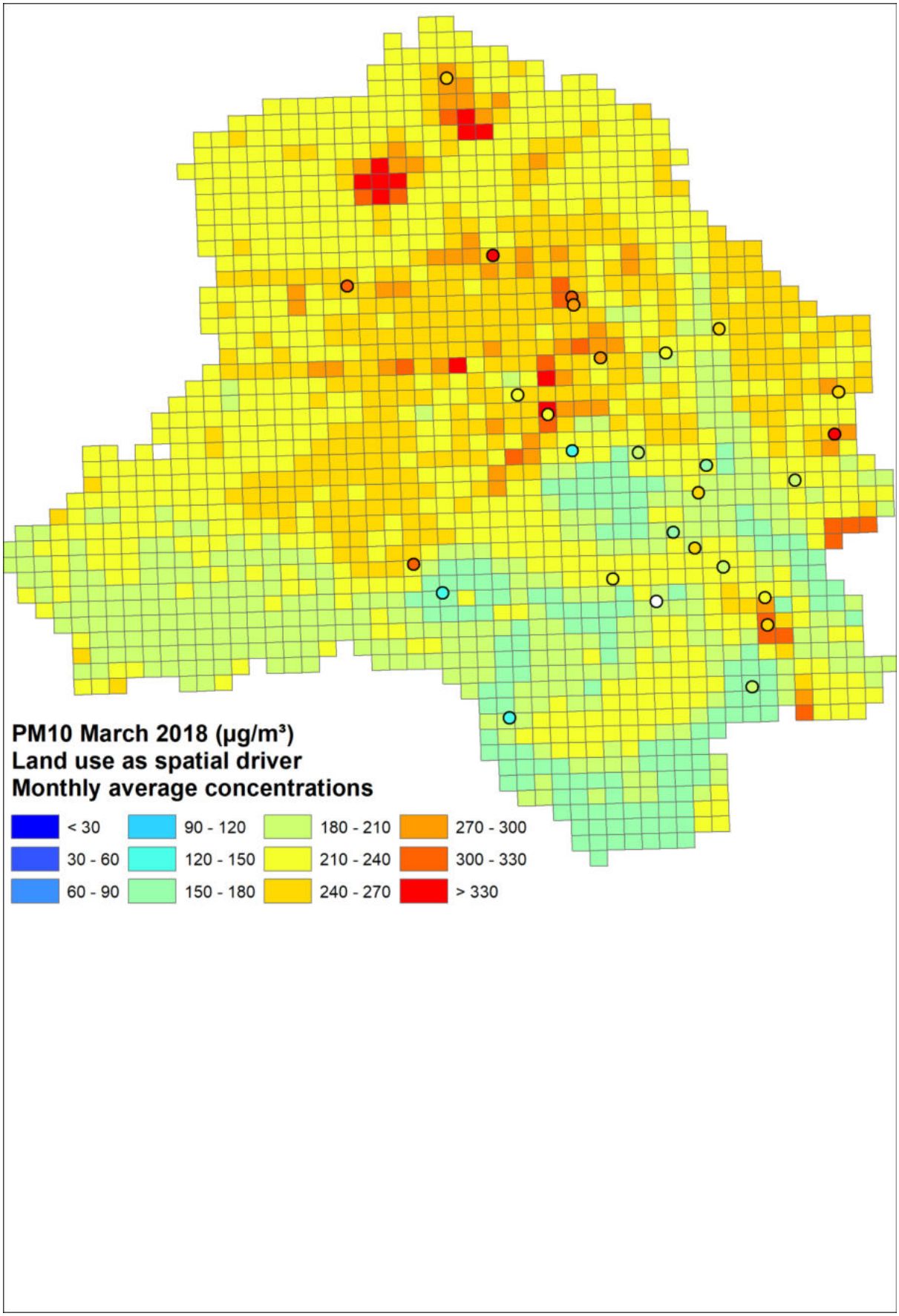

PM2.5 January 2019 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

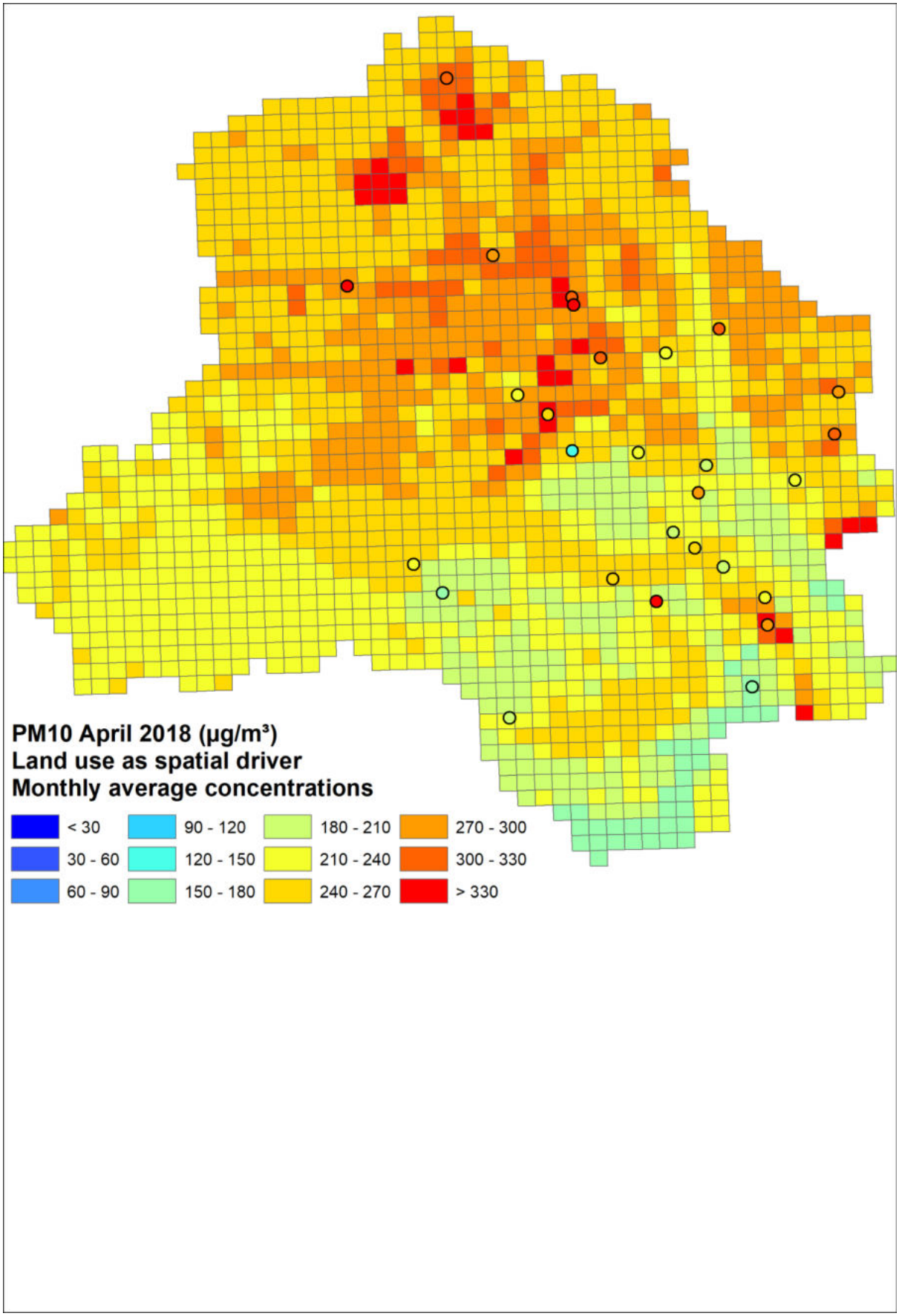

PM2.5 February 2019 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

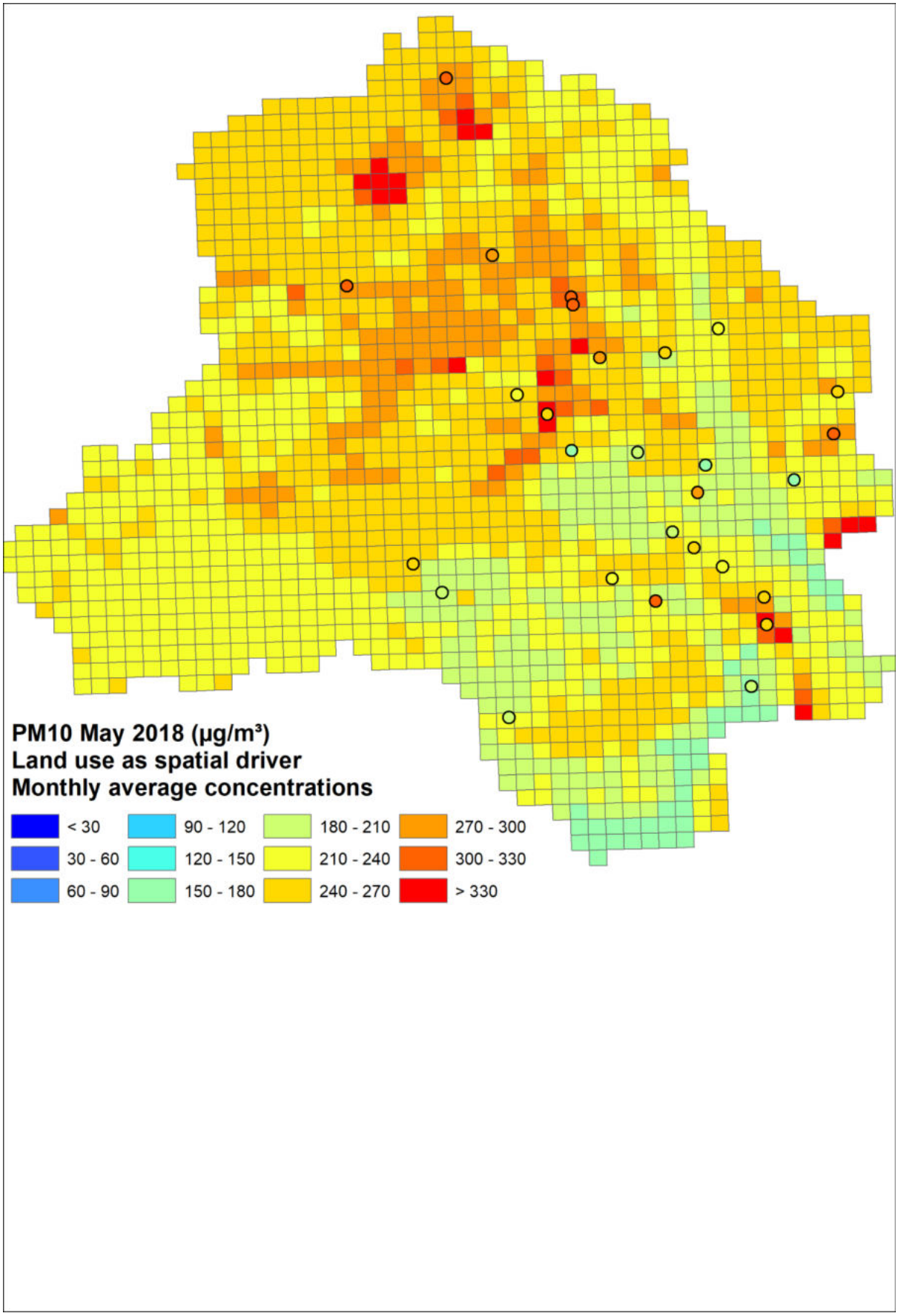

PM2.5 March 2019 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
> 240

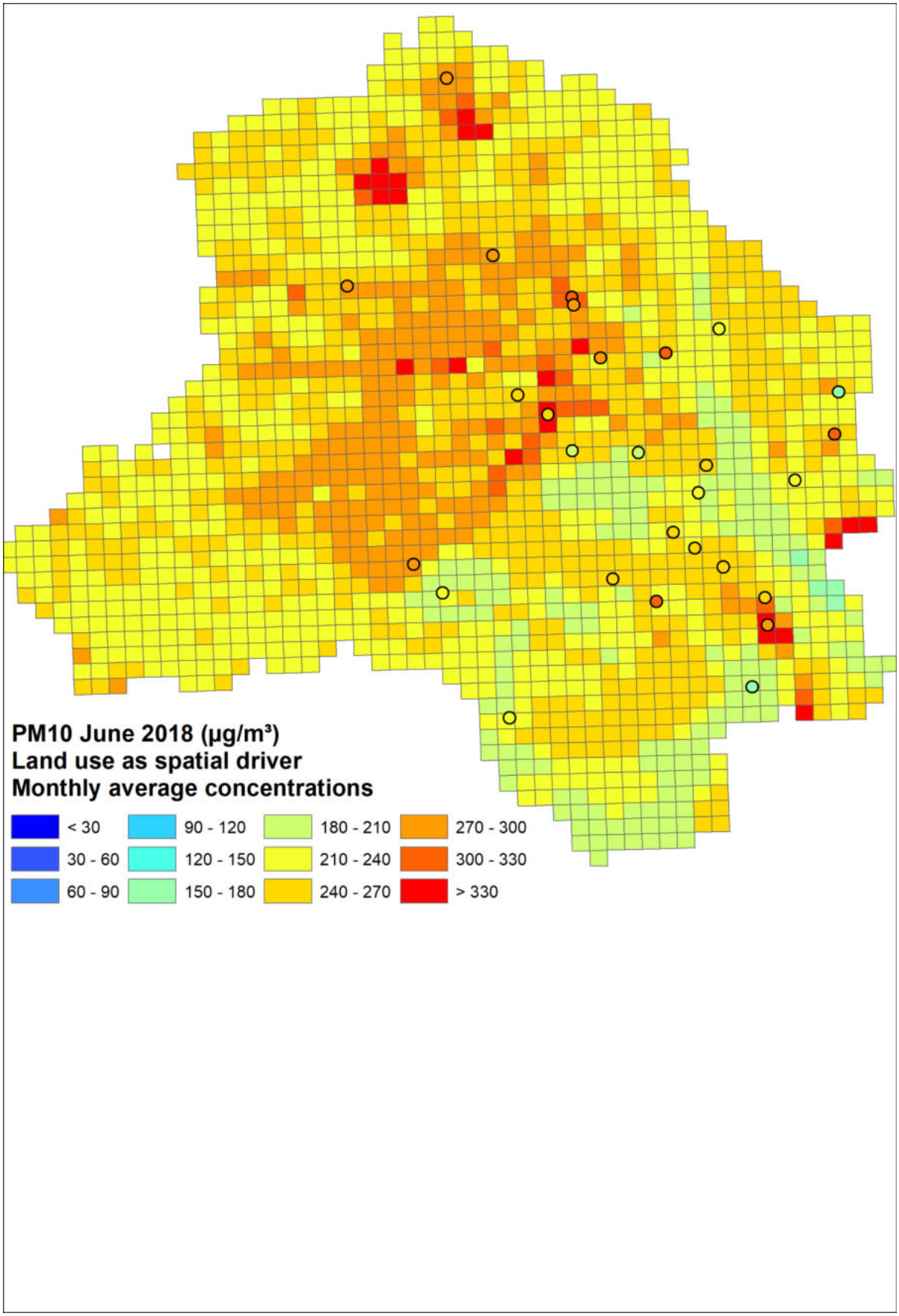

PM10 January 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

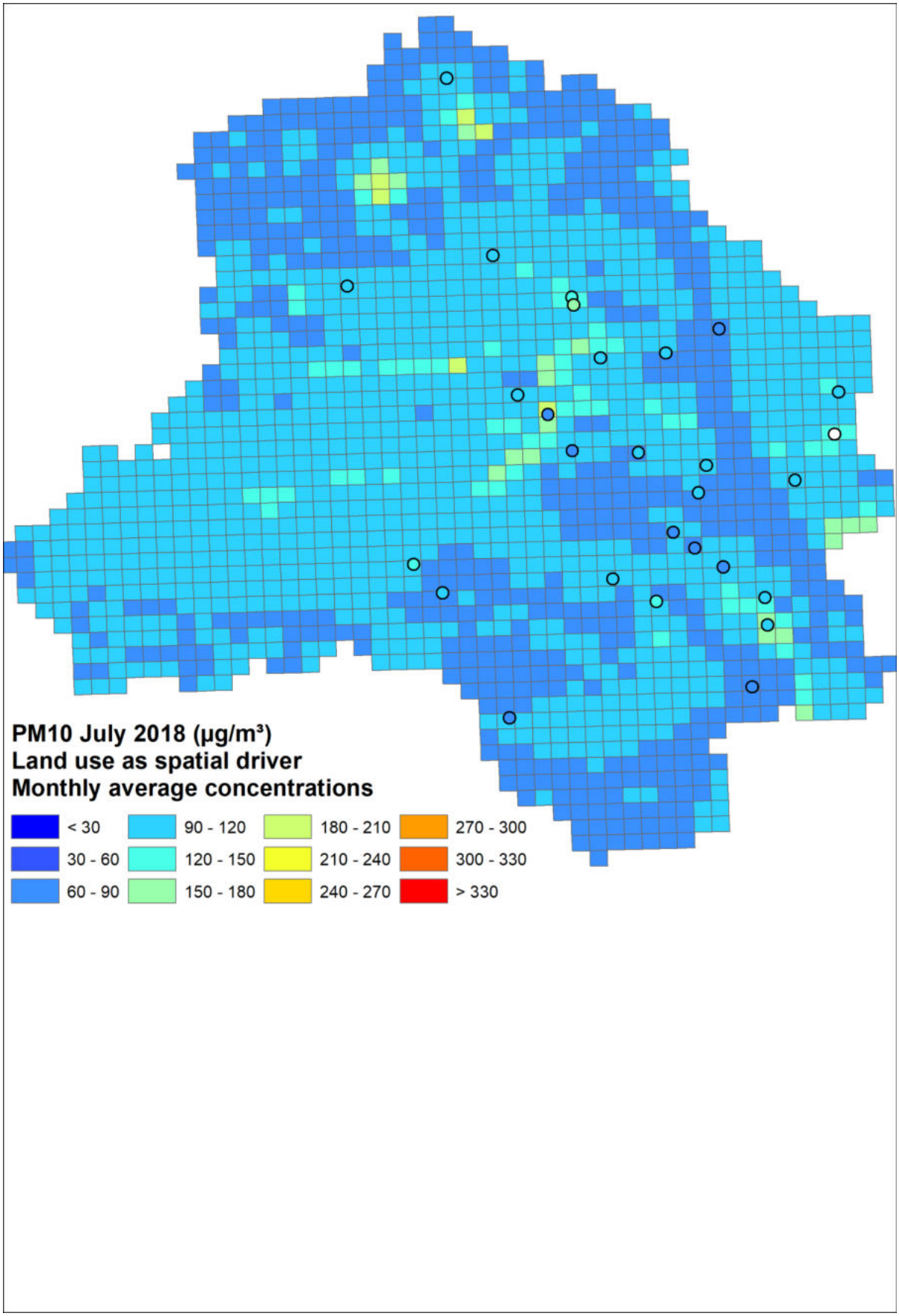

PM10 February 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

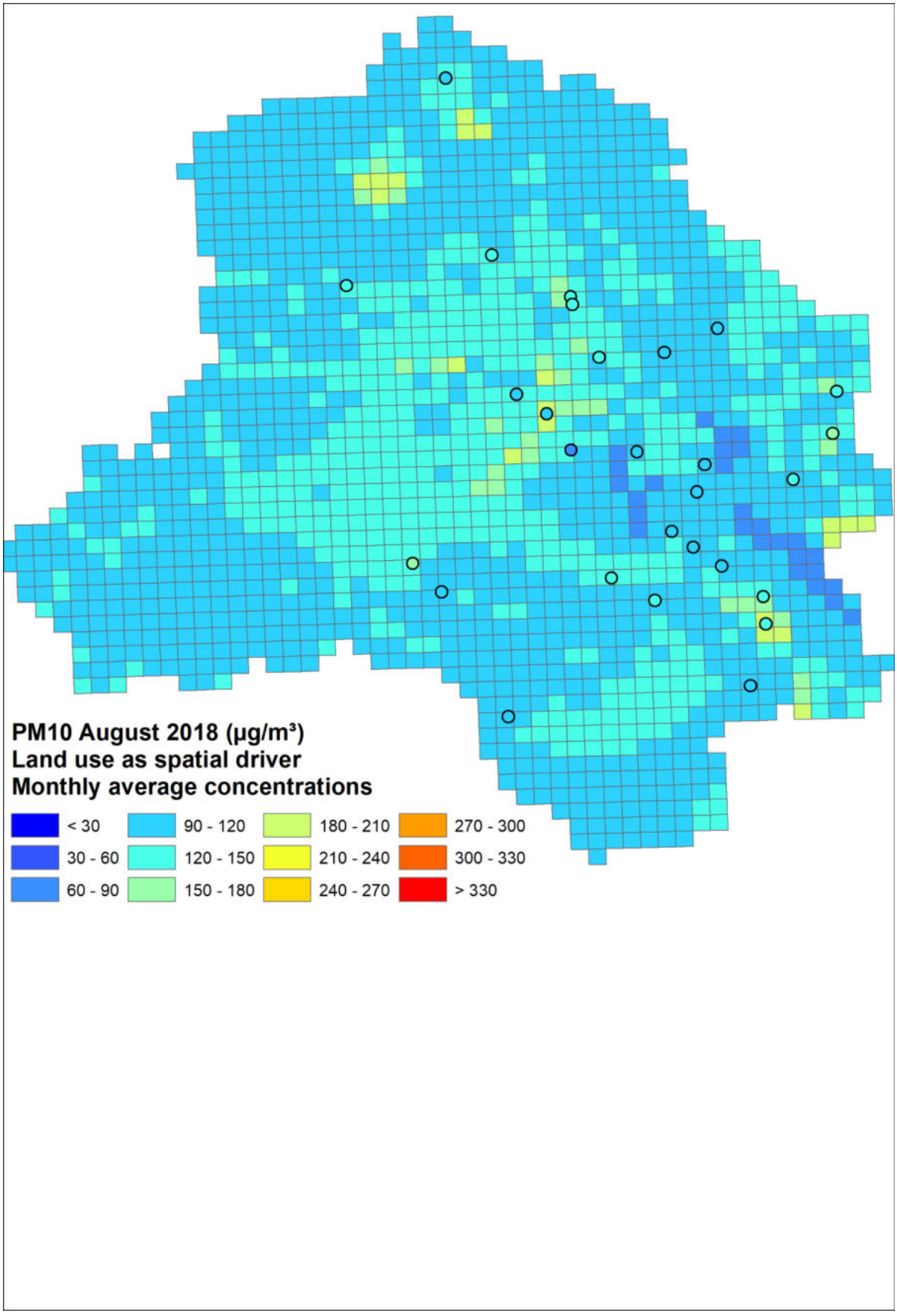

PM10 March 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

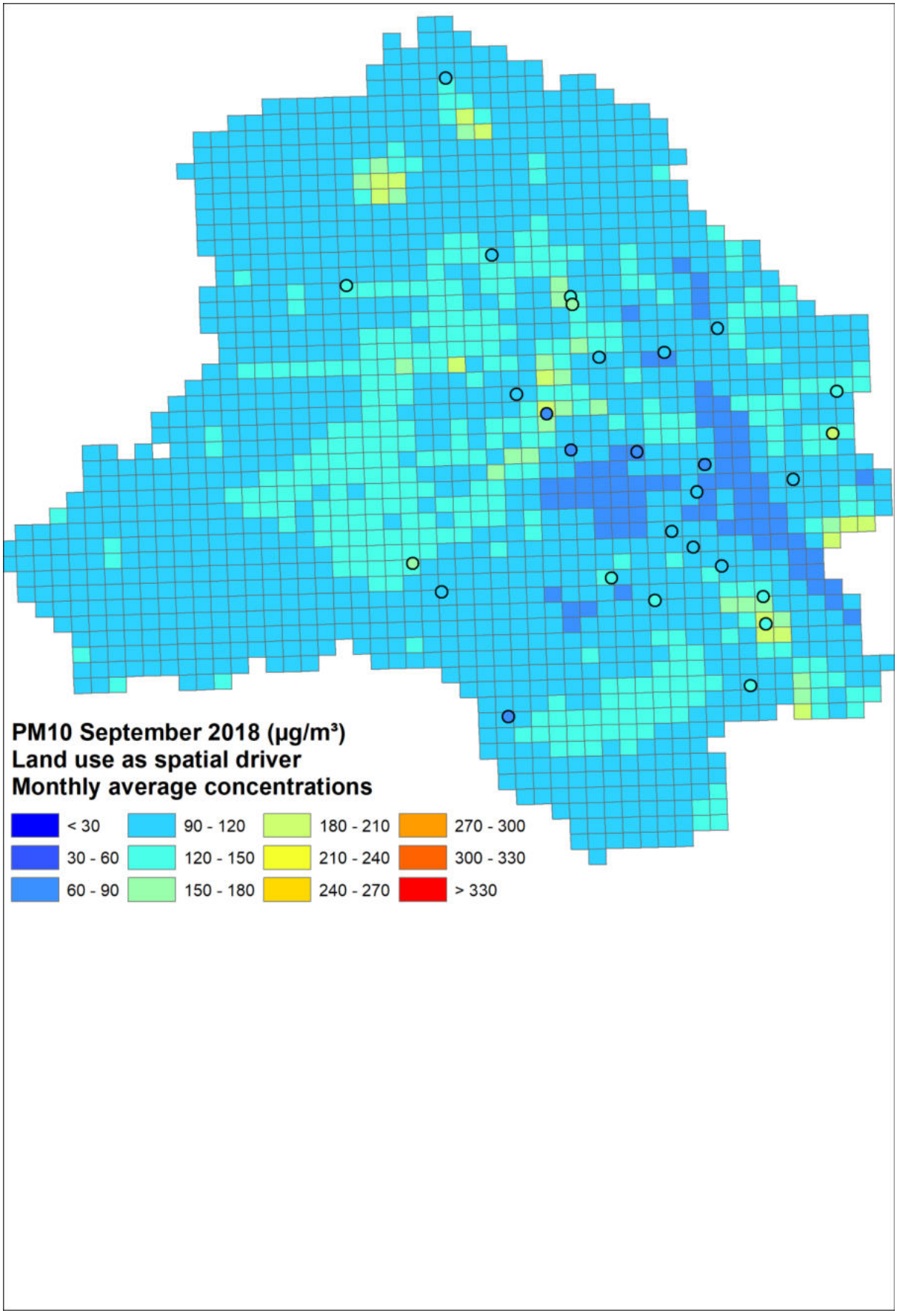

PM10 April 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

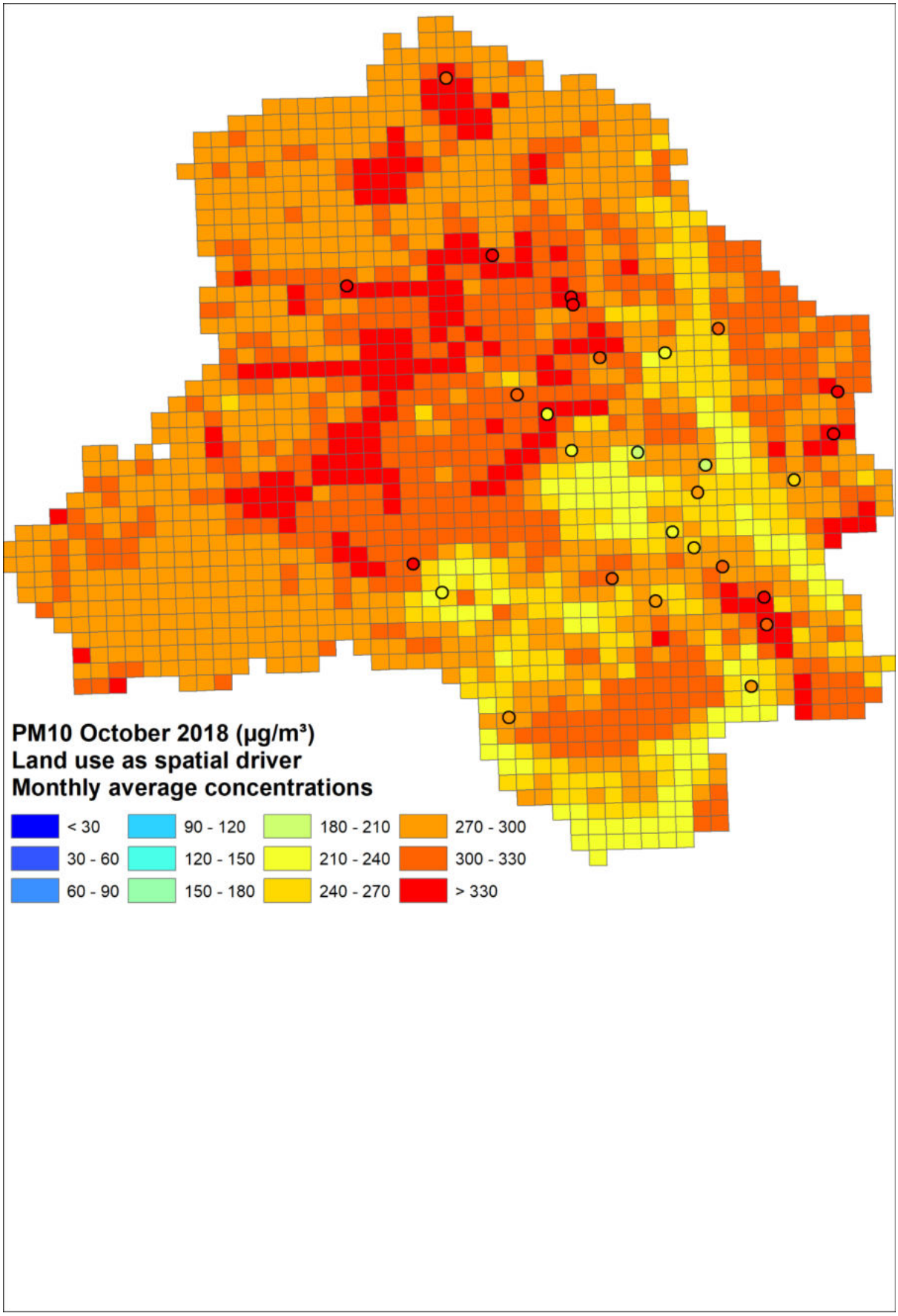

PM10 May 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

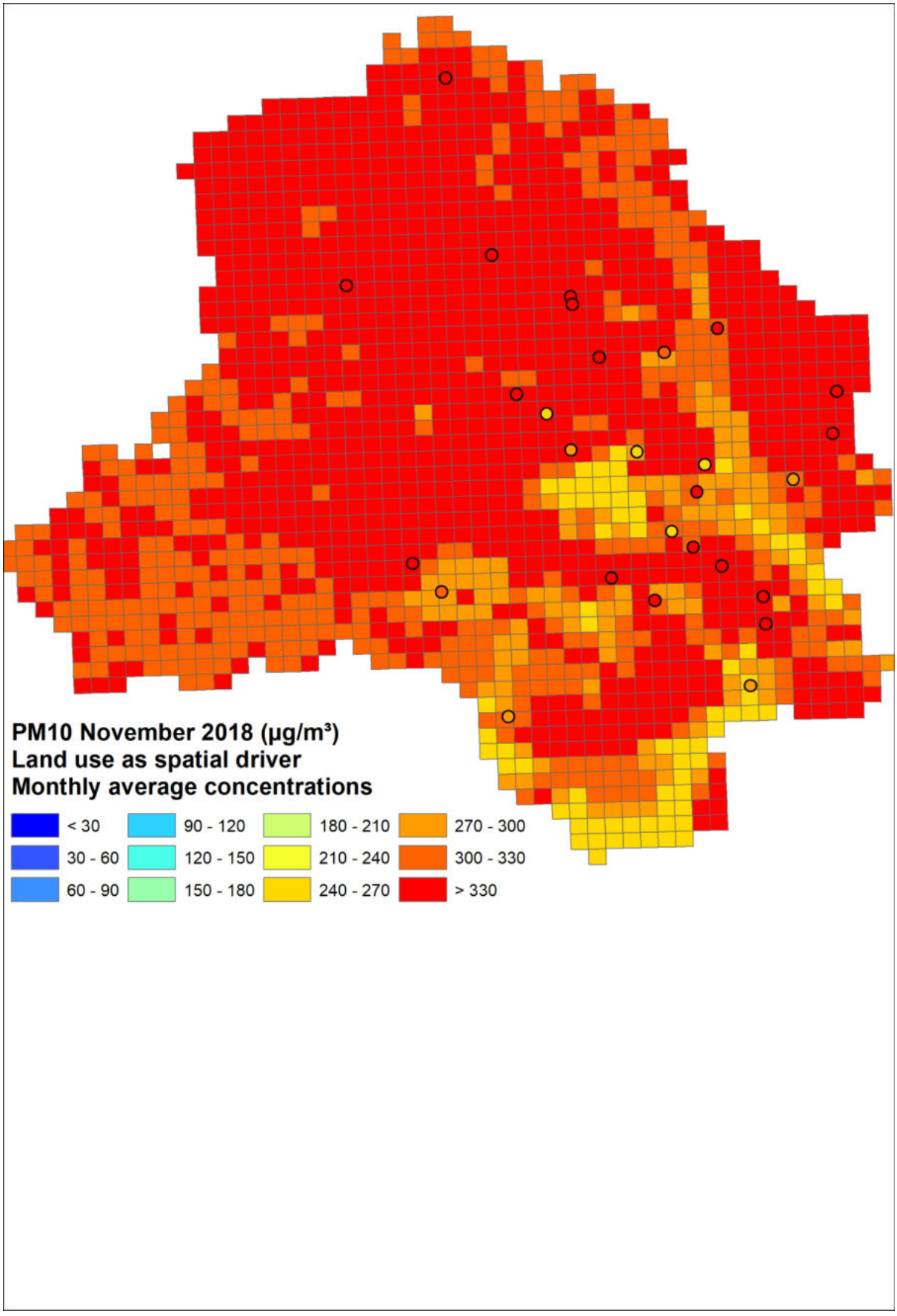

PM10 June 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

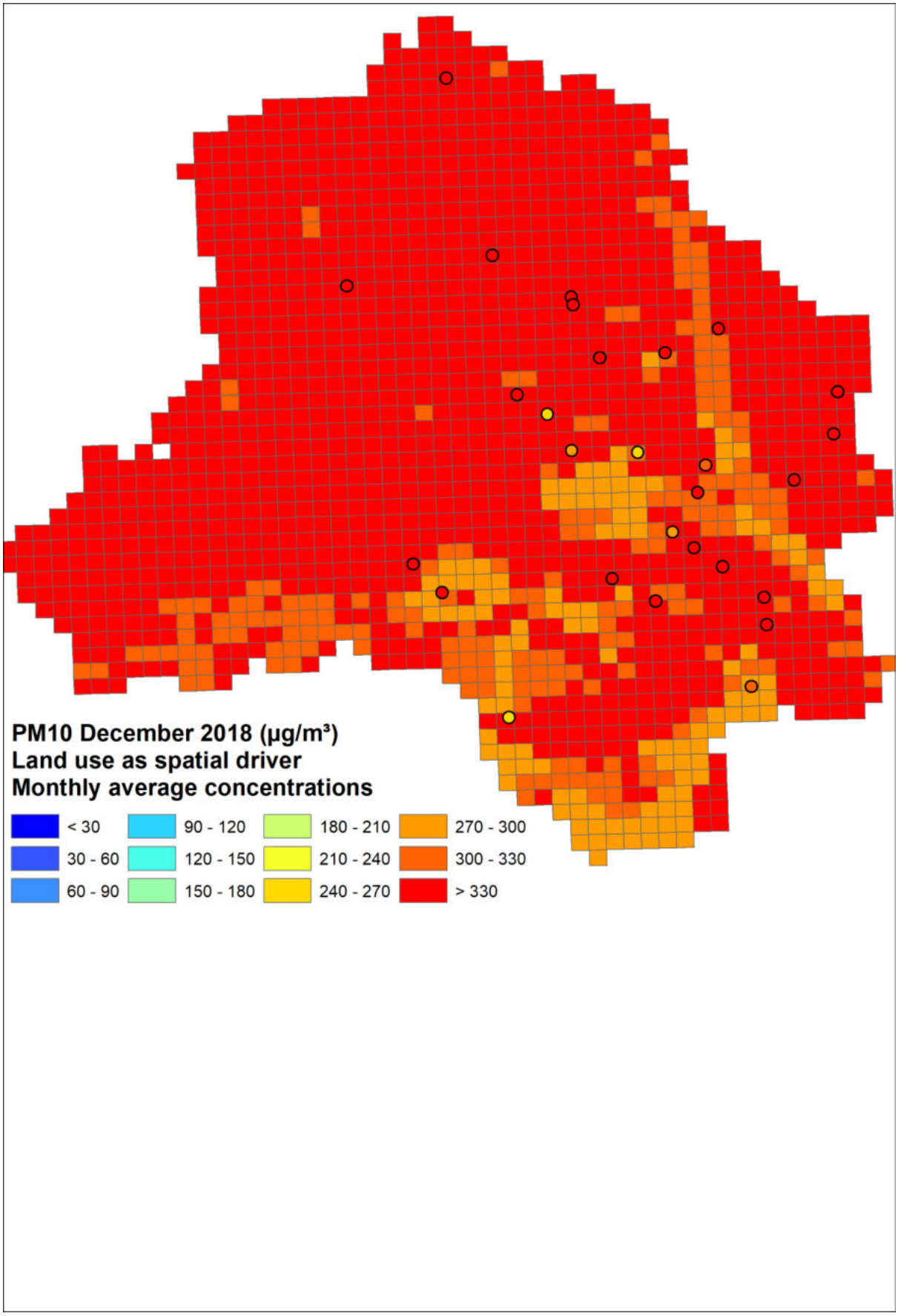

PM10 July 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

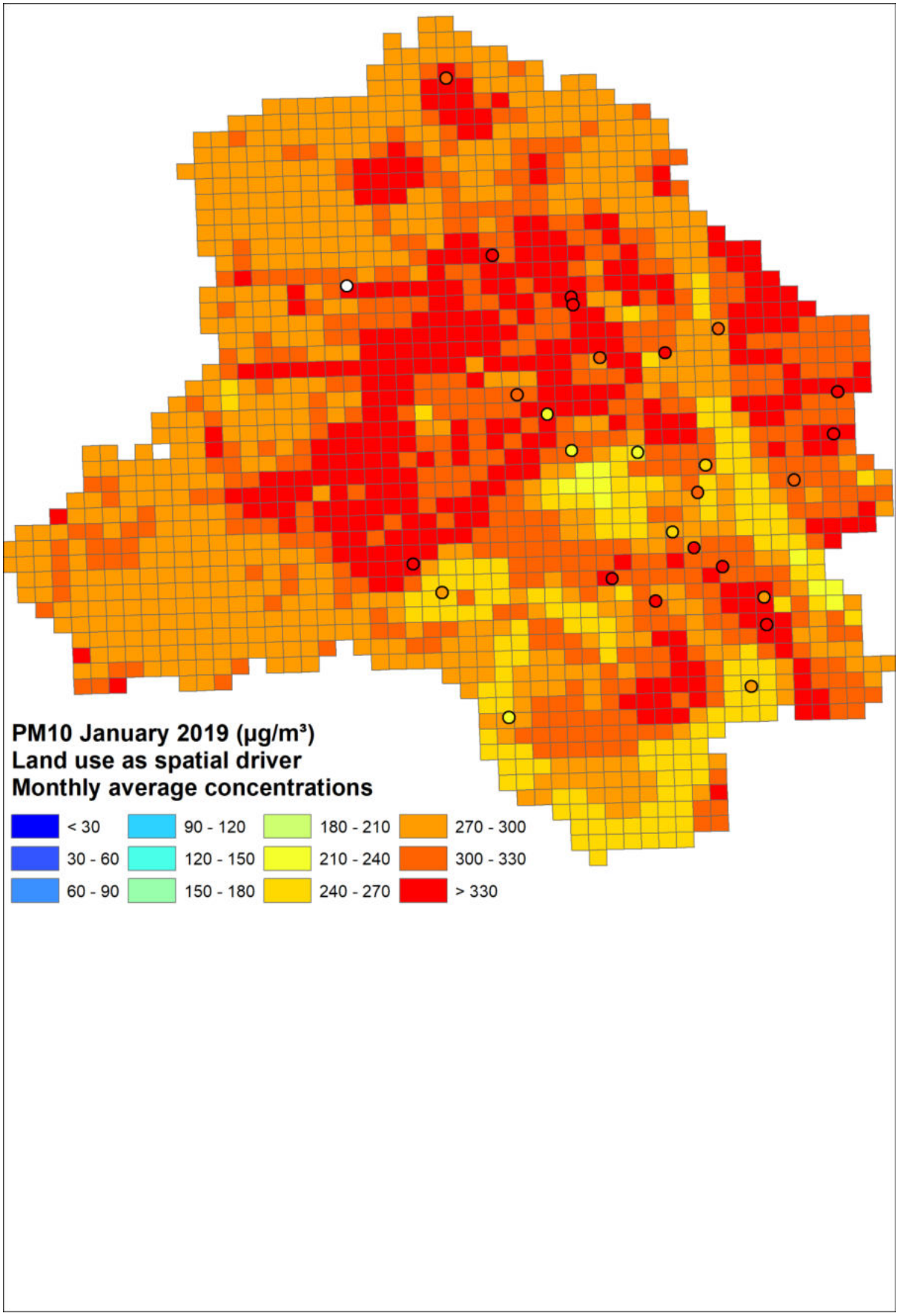

PM10 August 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

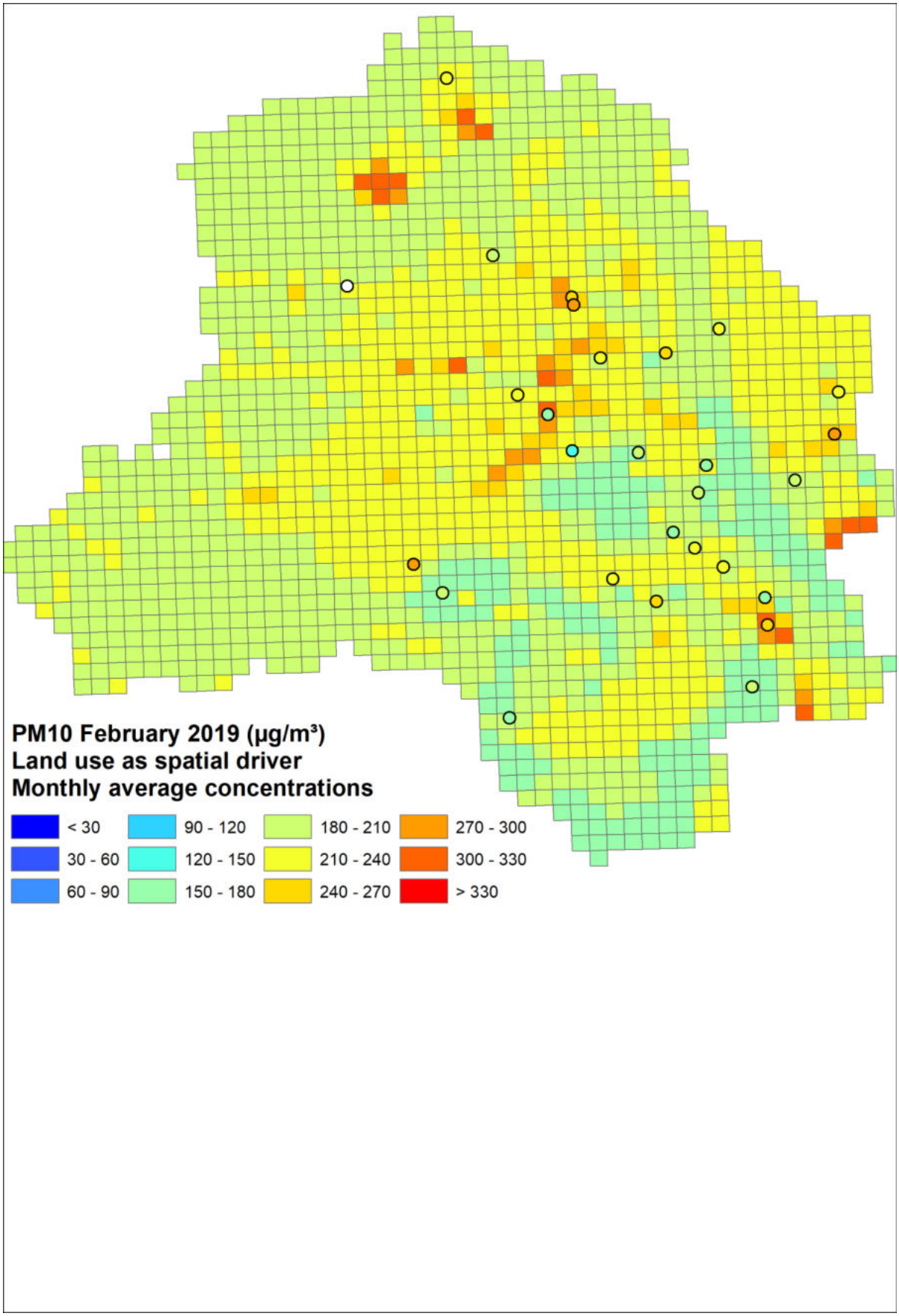

PM10 September 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

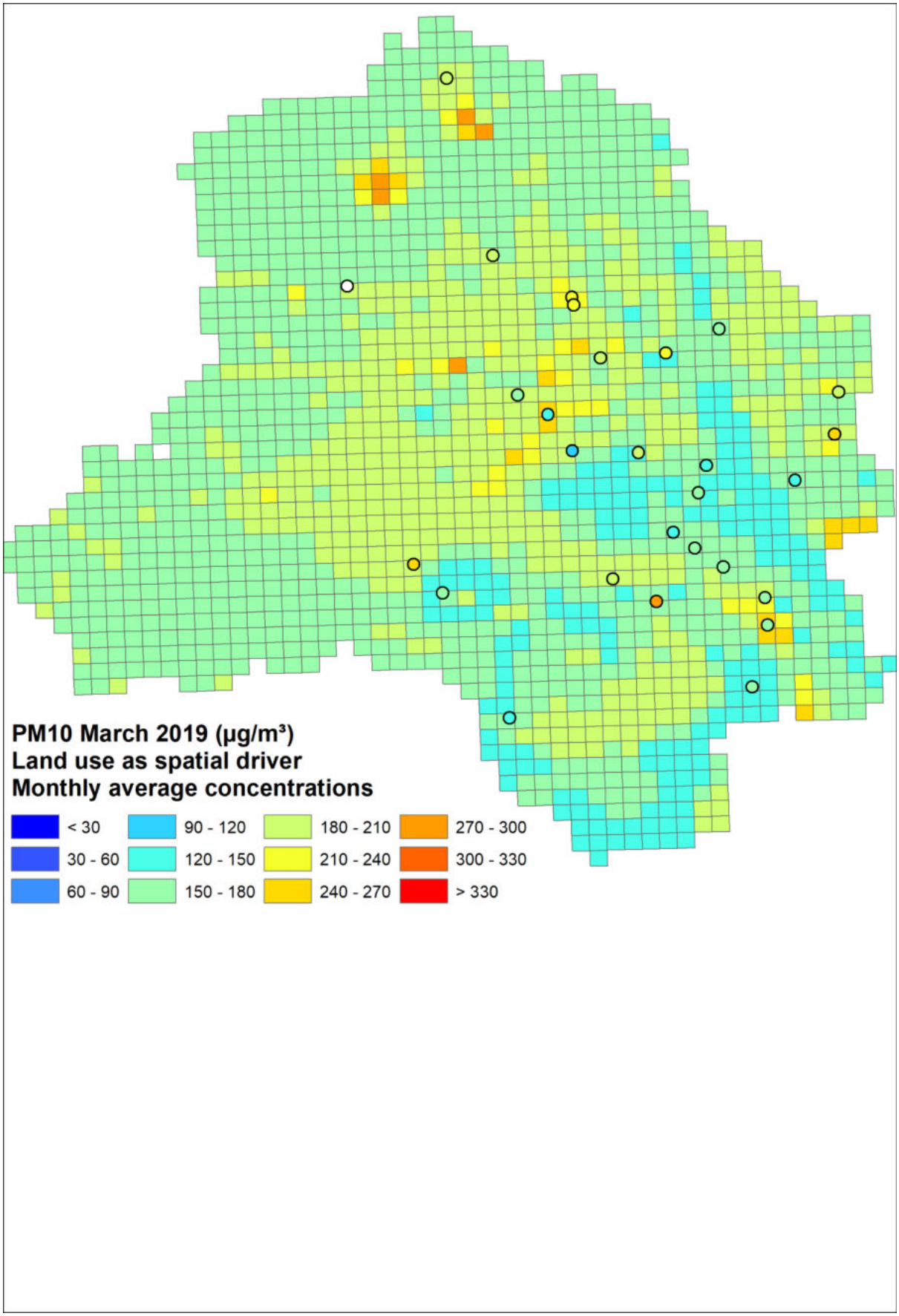

PM10 October 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330


PM10 November 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330


PM10 December 2018 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330


PM10 January 2019 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330


PM10 February 2019 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330


PM10 March 2019 (µg/m³)
Land use as spatial driver
Monthly average concentrations
< 30
30 - 60
60 - 90
90 - 120
120 - 150
150 - 180
180 - 210
210 - 240
240 - 270
270 - 300
300 - 330
> 330

Annexure II: Data cleaning procedures

Goal: to clean up the time series of all monitoring stations to have data suited to train the neural network modelling chain OPAQ by inspection of the data and filtering of spikes and insufficient data.

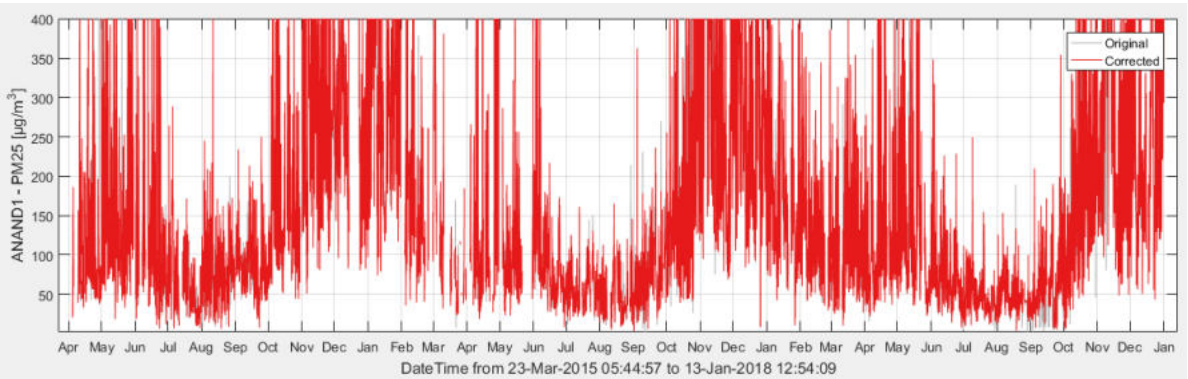
Tools:

- Filtering on hard limits: lower limit and upper limit (example: stations reporting incorrect data such as NO_2 and PM levels below 1)
- Removing a selection of dates (example: stations reporting constant data for several days)
- Hampel filter and Gaussian anomaly detection to remove spikes
- Removing some of the stations as the data show a strange behaviour

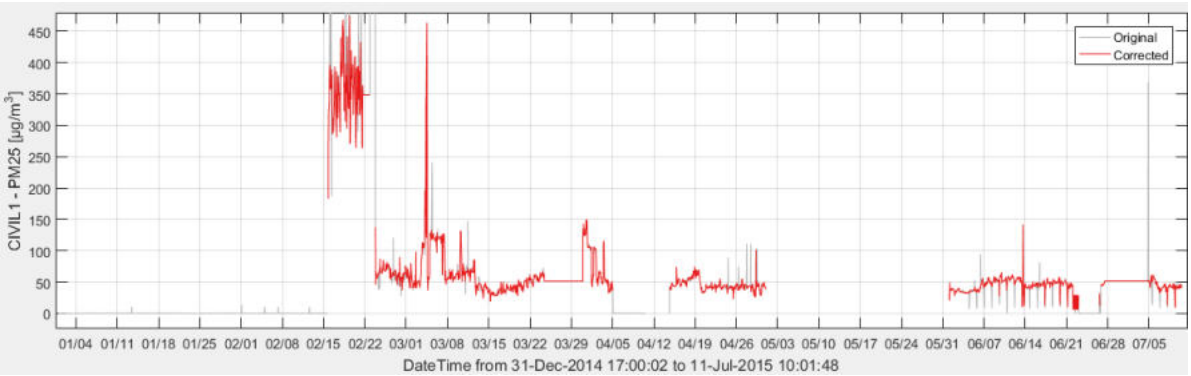
The following steps have been taken to clean all data:

- For each station, the time series of the observations have been plotted and visually inspected to leave out (part of) time series which show strange trends (for example constant observations over longer periods)
- A clean up step has been completed, to ensure we are only using valid observations. This consist of two step. Firstly a hard limit filter has been applied to exclude observations which are lower than 2 µg/m³ for PM_{10} and $PM_{2.5}$, 1 µg/m³ for NO_2, which seems to be 'no data' but registered as '0' or '1'. This output is used for RIO to generate air quality maps for Delhi. Secondly, I have applied a Hampel filter (logtransform, window width 3, threshold 3) to remove spikes. For each sample of x, the function computes the median of a window composed of the sample and its six surrounding samples, three per side. It also estimates the standard deviation of each sample about its window median using the median absolute deviation. If a sample differs from the median by more than three standard deviations, it is replaced with the median. The filter is applied on log-transformed values. *More information Hampel filter, see* https://nl.mathworks.com/help/signal/ref/hampel.html.

For example, the clean-up process carried out for $PM_{2.5}$ time series of station Anand Vihar and Civil Lines is shown in figure below. Similar approach has been carried out for all the stations for pollutants PM_{10}, $PM_{2.5}$, O_3 and NO_2.



$PM_{2.5}$ time series of station Anand Vihar, original in grey, corrected data in red. Data have been used after filtering.

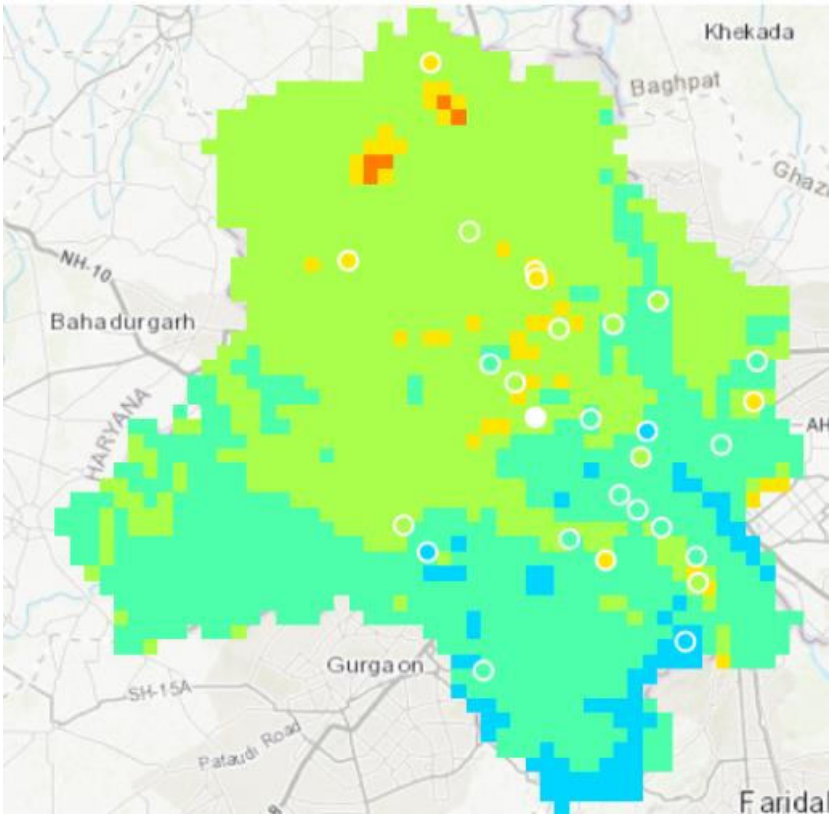


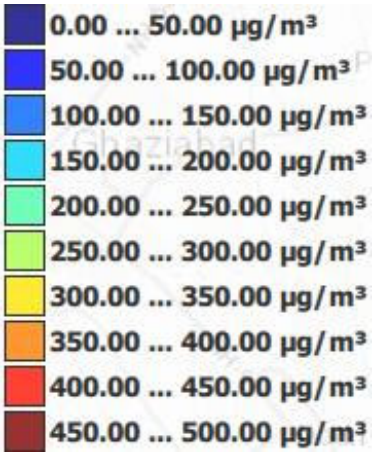
$PM_{2.5}$ time series of station Civil Lines, data have NOT been used due to inadequate data points.

Annexure III: Seasonal Spatial Map for hotspot identification

Spatial PM_{10} maps for different seasons: Summer, post-monsoon and winter is represented are made below. The spatial map clearly suggests RIO interpolation technique has able to introduce land used based local variation in estimating PM_{10} concentration and identifying the potential hotspots locations across entire territory of Delhi. The concentration during different seasons followed seasonal trend, predicting higher values during winter and post-monsoon months and least during summer period. The model predicted concentration in the range 250-300 µg/m³ in northern and central part of domain region and fewer locations estimated with higher concentration (400-450 µg/m³) mostly characterized by urban and industrial land cover pattern. However, the hot spot regions with estimated concentration in the range 350-400 µg/m³ identified during post-monsoon and winter seasons were located in north-western, central and few south-eastern part of city domain. The hot spots identified during post-monsoon and winter phase were mostly characterized by residential, industrial and traffic bound locations indicating local emissions could have phenomenon impact on regional air quality. It was observed a significant proportion of locations mostly characterized by industrial land cover were identified as hot spot locations suggesting the concentration at such location remained higher throughout the year regardless of season. In fact it is also observed from the maps regardless of season the model still envisaged with highest concentration (400-450 µg/m³) in north-western part of city domain. Higher concentration estimated during winter period could be because of favourable meteorological conditions: temperature inversion, low wind patterns and shallow boundary layer heights.

SUMMER (April-May-June)

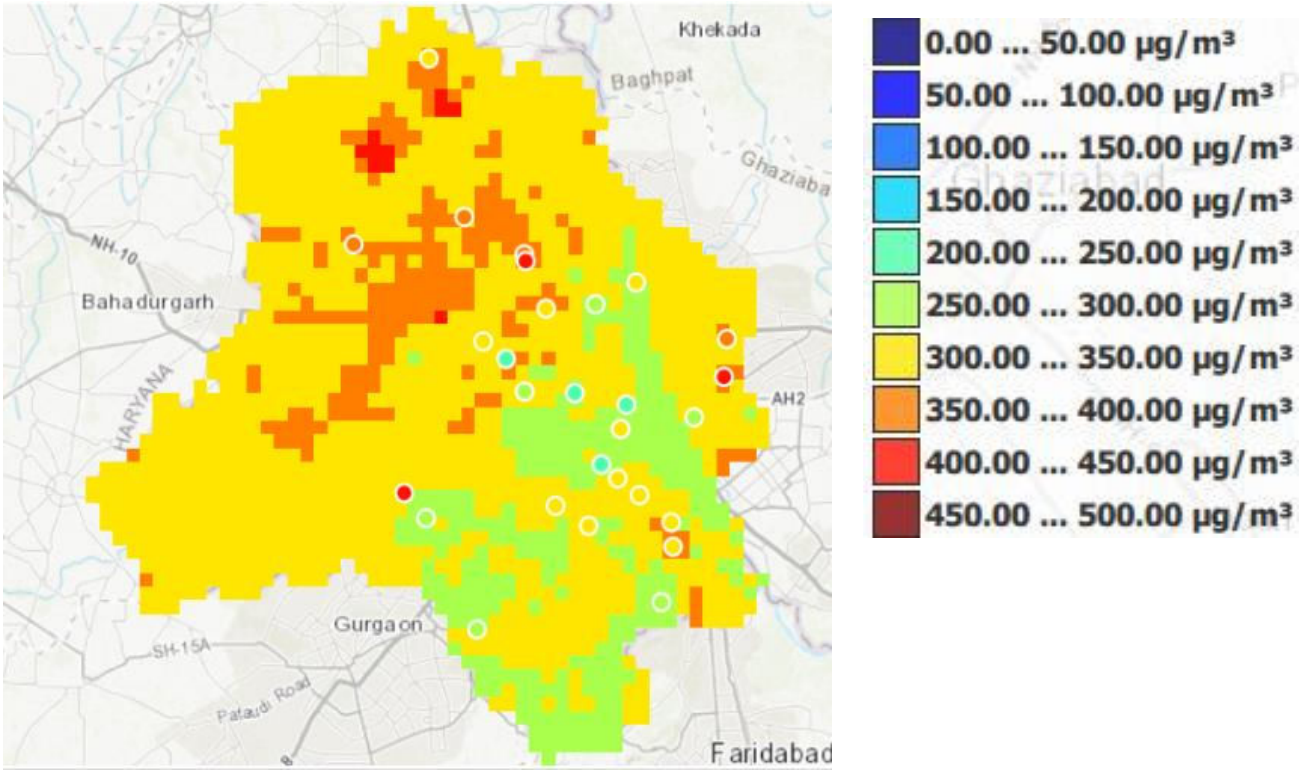




1) BAWANA
2) SHAHPUR GARHI VILLAGE-RAZA PUR VILLAGE-BHOR GARH VILLAGE-KURENAI VILLAGE
3) KANJHAWALA VILLAGE
4) SWEEPERS COLONY-SECTOR 18F
5) SANJAY NAGAR-JJ COLONY-POORVI BLOCK BB-SHALIMAR BAGH INDUSTRIAL AREA-BLOCK J4
6) JJ COLONY 2-BLOCK A
7) PEERAGARHI UDYOG NAGAR
8) SHARDA NIKETAN
9) DR LOHIA-RAMPURA VILLAGE-TRI NGAR-SHAKTI NAGAR
10) DILKHUSH BAGH INDUSTRIAL AREA
11) ISHWAR COLONY-SATYAWATI NAGAR
12) NAYI BASTI-THAKABAPA NAGAR
13) BALJIT NAGAR

14) MAYAPURI INDUSTRIAL AREA PHASE 1
15) OKHLA INDUSTRIAL AREA

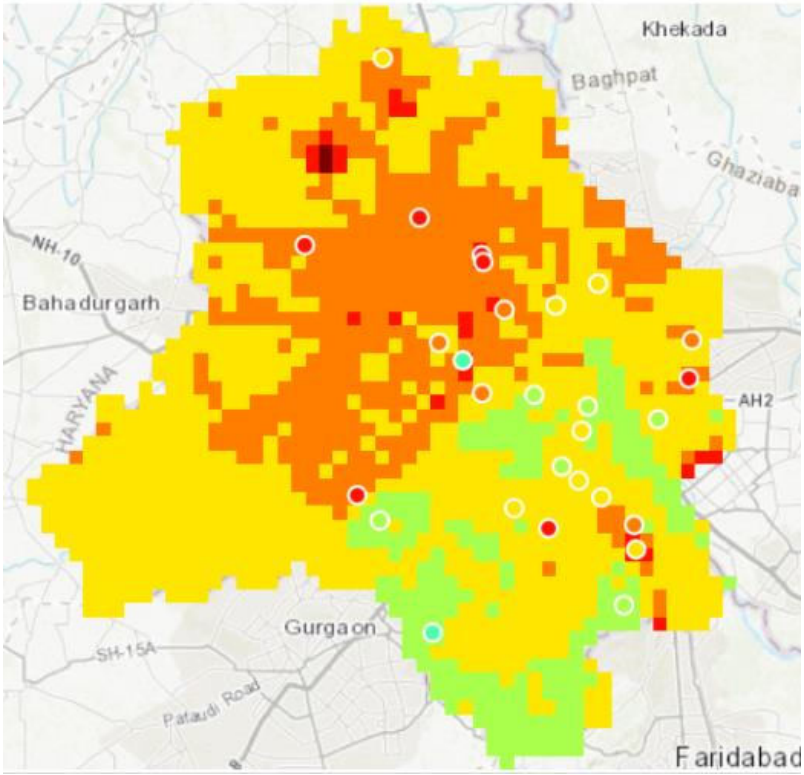
POST MONSOON (Oct-Nov)

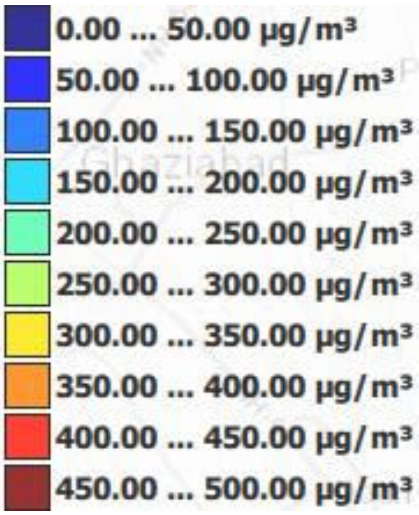


1) BAWANA
2) SHAHPUR GARHI VILLAGE-RAZA PUR VILLAGE-BHOR GARH VILLAGE-KURENAI VILLAGE
3) MANGOLPURI INDUSTRIAL AREA-NEW MULTAN NAGAR
4) OKHLA INDUSTRIAL AREA
5) LOTUS TEMPLE-KALKAJI MANDIR-NEHRU PLACE
6) UDYOG MARG-DALLUPURA MARG
7) DR BARMAN MARG
8) VISHWAKARMA NAGAR
9) DR LOHIA-RAMPURA VILLAGE-TRI NGAR-SHAKTI NAGAR
10) ASHOK VIHAR POLICE STATION ROAD
11) DILKHUSH BAGH INDUSTRIAL AREA
12) STYAWATI COLLEGE
13) DR. HEDGEWAR MARG
14) PEERAGARHI UDYOG NAGAR
15) RAMA ROAD INDUSTRIAL AREA-NAJAFGARH ROAD
16) NAYI BASTI-THAKABAPA NAGAR
17) SHALIMAR BAGH INDUSTRIAL AREA

18) BADLI INDUSTRIAL AREA
19) MAYAPURI INDUSTRIAL AREA PHASE 1
20) JHILMIL METRO STATION
21) JHARODA MAJRA BURARI
22) JHANGIR PURI VILLAGE
23) KADI VIHAR
24) ROSHAN PURA-ROSHAN GARDEN-NAJAFGARH NANGLO ROAD

WINTER (Dec-Jan-Feb)



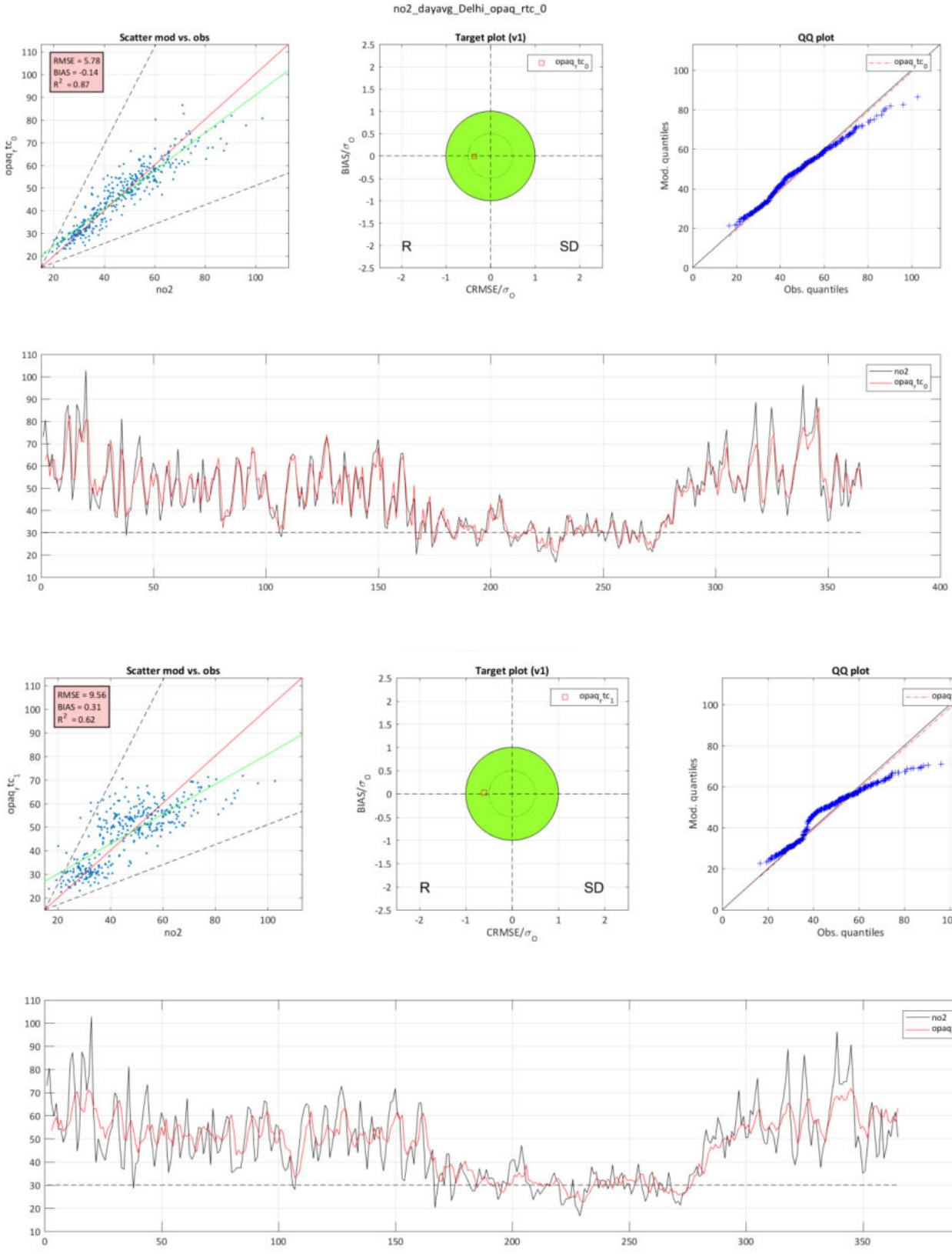


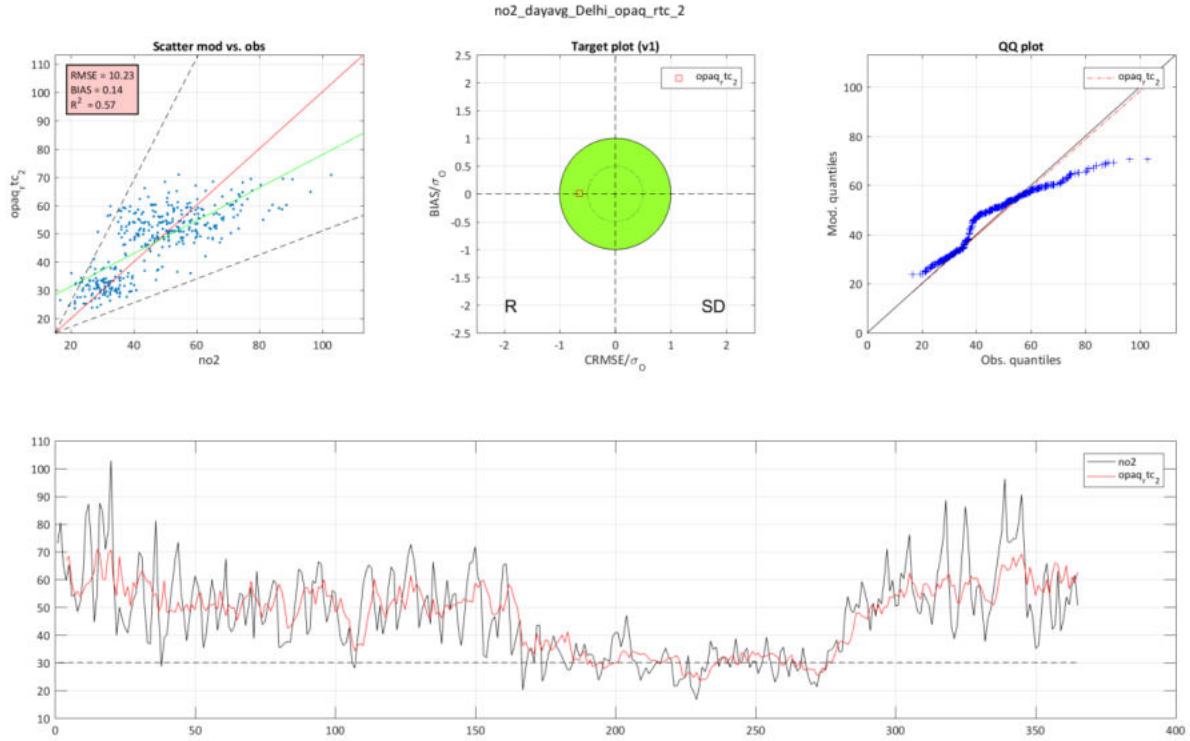
1) BAWANA
2) SHAHPUR GARHI VILLAGE-RAZA PUR VILLAGE-BHOR GARH VILLAGE-KURENAI VILLAGE
3) SOM BAZAR ROAD-BEGAMPUR VILLAGE-PEHLADPUR ROAD
4) BADLI INDUSTRIAL AREA-SAHIBABAD DAIRY BLOCK B
5) SHALIMAR BAGH INDUSTRIAL AREA
6) SANJAY NAGAR
7) DR. KN KATJU MARG ROHINI
8) DR LOHIA-RAMPURA VILLAGE-TRI NGAR-SHAKTI NAGAR
9) ASHOK VIHAR POLICE STATION ROAD
10) DR. HEDGEWAR MARG-NEW MULTAN NAGAR BLOCK E
11) JJ COLONY 1 KAVITA COLONY-NANGLOI SULATANPURI ROAD

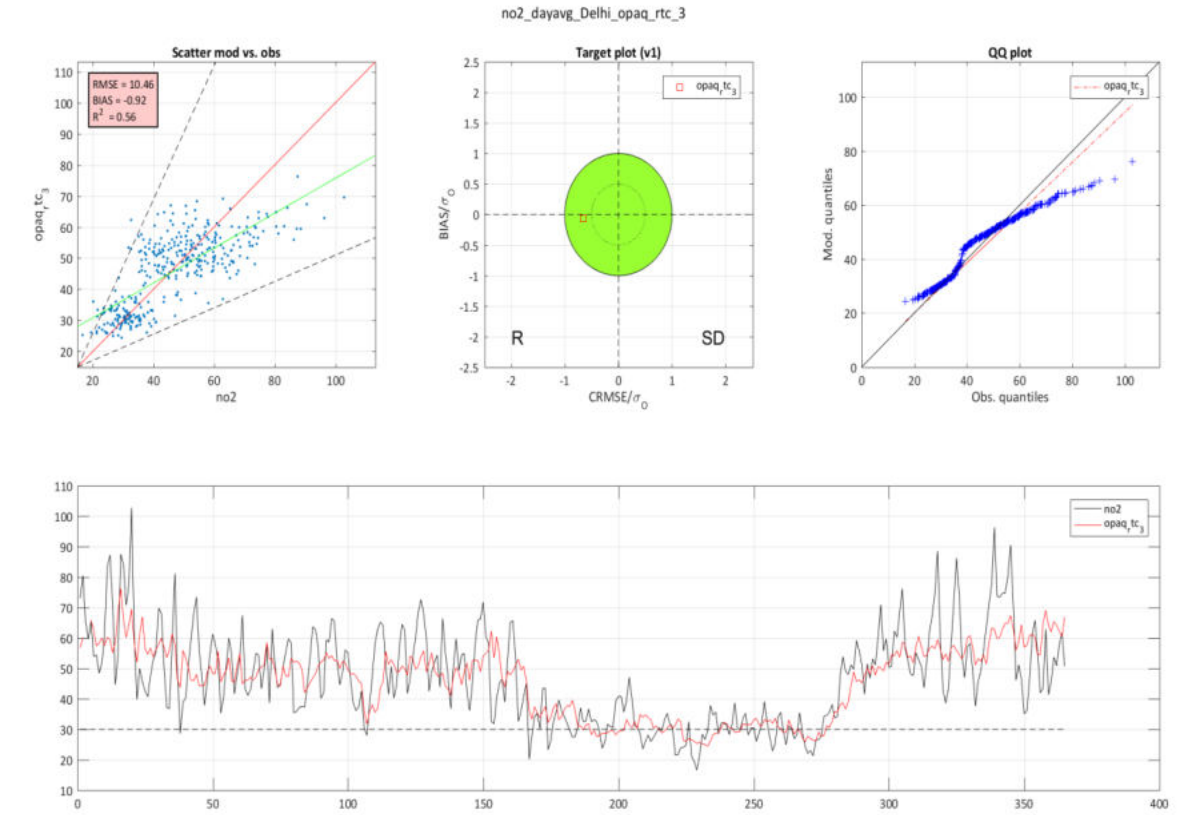
12) BALJIT NAGAR
13) RAMA ROAD INDUSTRIAL AREA-NAJAFGARH ROAD
14) DR BARMAN MARG-MADHU VIHAR
15) INDANE GAS FACTORY
16) OKHLA INDUSTRIAL AREA
17) LOTUS TEMPLE-KALKAJI MANDIR-NEHRU PLACE
18) JHARODA MAJRA BURARI
19) JHANGIR PURI VILLAGE-WEST SANT NAGAR

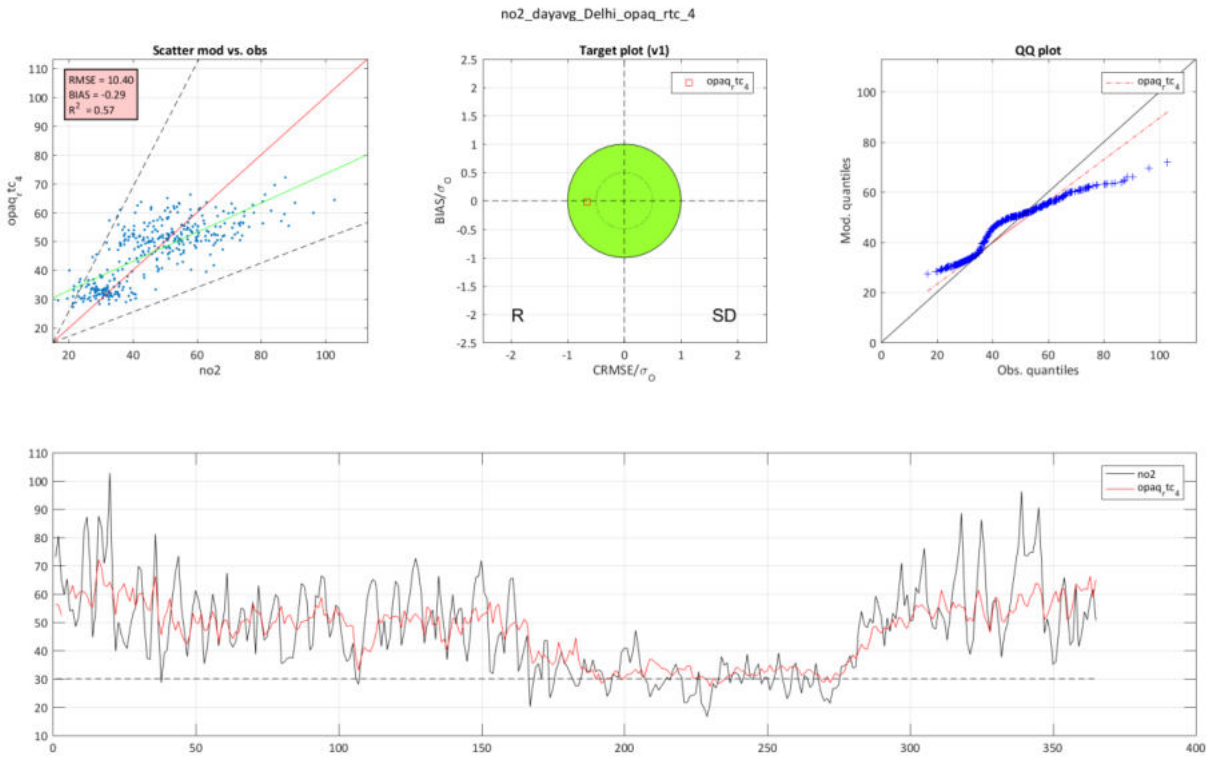
Annexure IV: Forecast for 2019 PM_{10}, $PM_{2.5}$, O_3, NO_2

NO_2 (Arranged from Day0 to Day4)

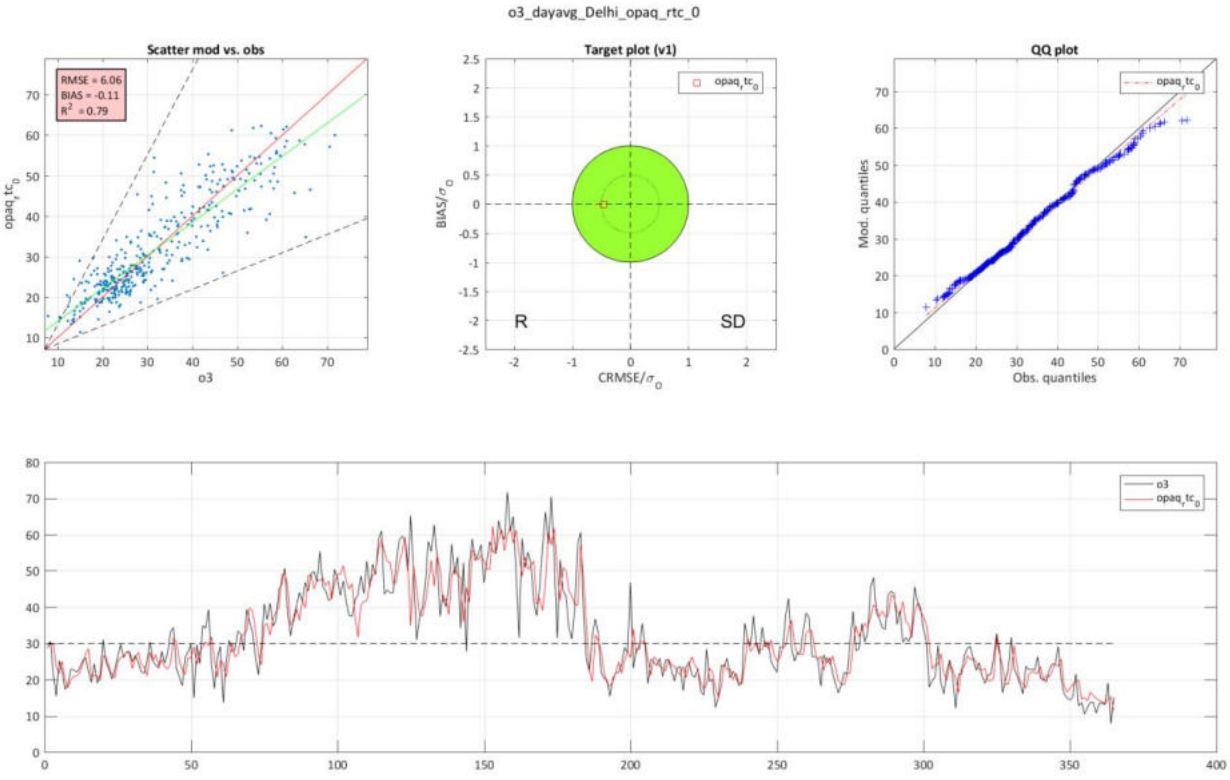


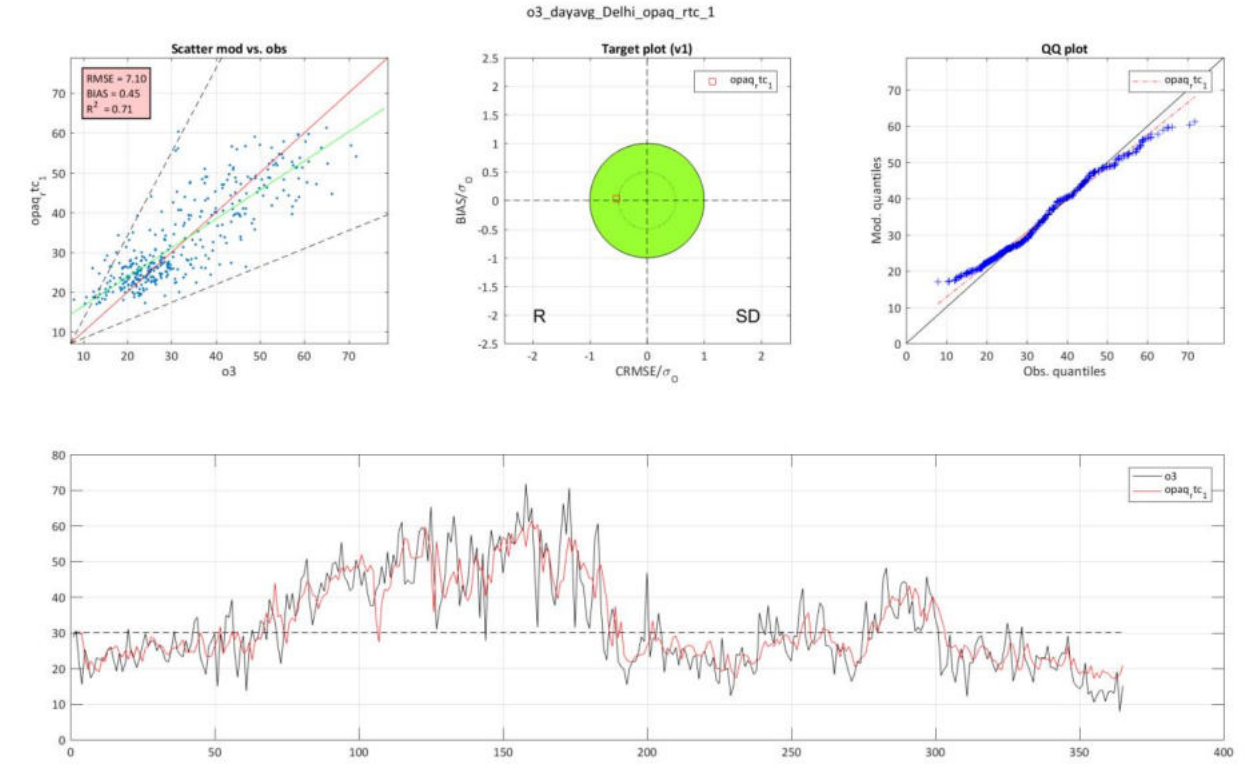

no2_dayavg_Delhi_opaq_rtc_2
Scatter mod vs. obs
RMSE = 10.23
BIAS = 0.14
R² = 0.57
opaq_rtc_2
no2
Target plot (v1)
BIAS/σ_O
CRMSE/σ_O
R
SD
QQ plot
Mod. quantiles
Obs. quantiles
no2
opaq_rtc_2

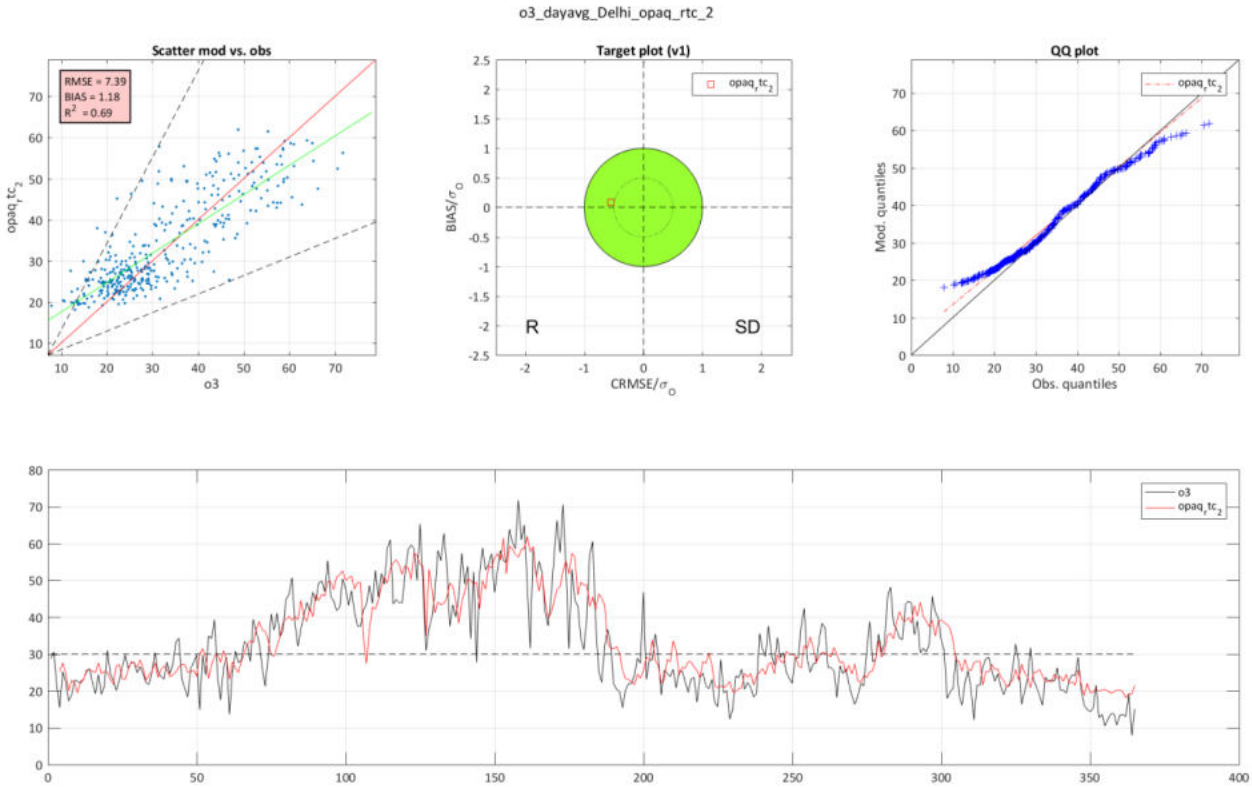

no2_dayavg_Delhi_opaq_rtc_3
Scatter mod vs. obs
RMSE = 10.46
BIAS = -0.92
R² = 0.56
opaq_rtc_3
no2
Target plot (v1)
BIAS/σ_O
CRMSE/σ_O
R
SD
QQ plot
Mod. quantiles
Obs. quantiles
no2
opaq_rtc_3

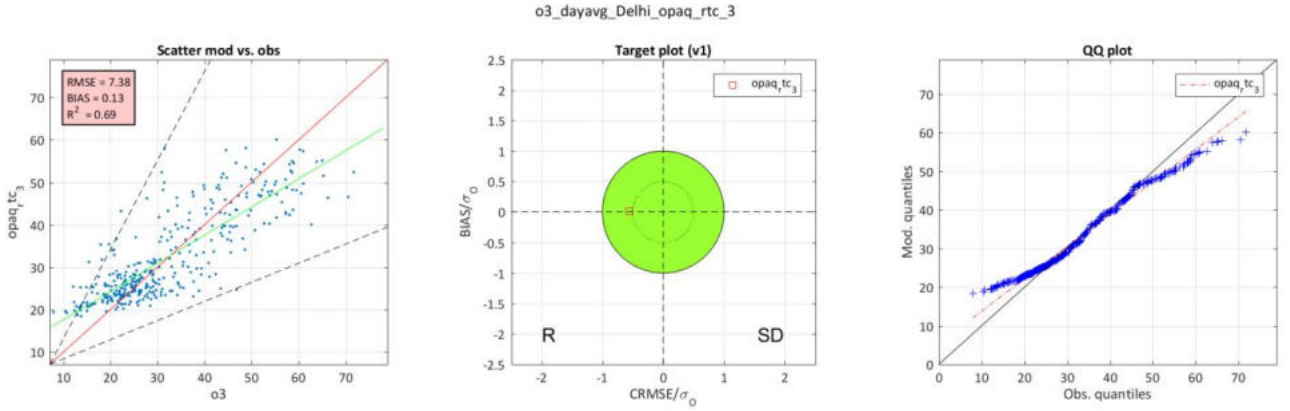


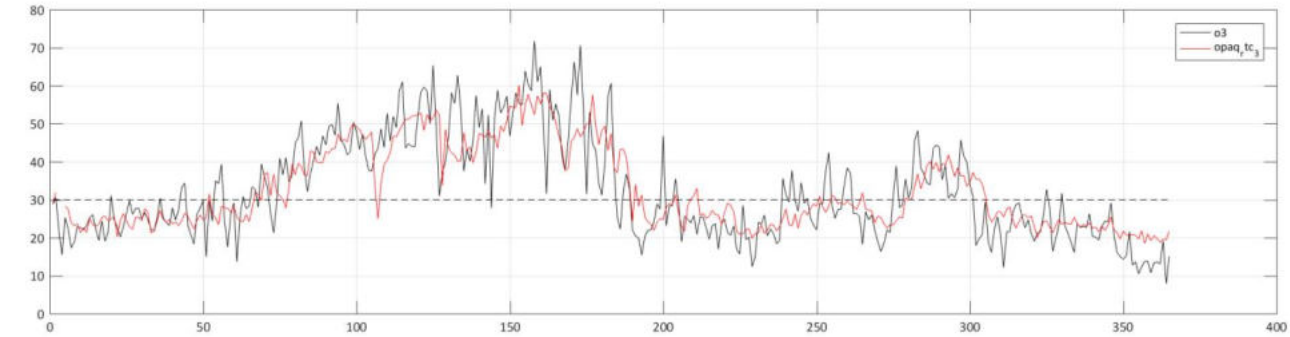
O_3 (Arranged from Day0 to Day4)

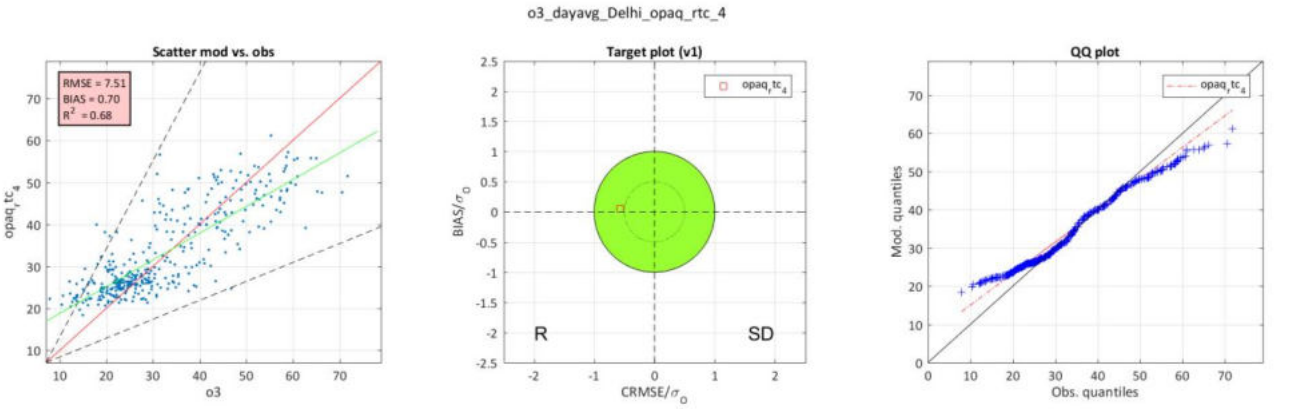


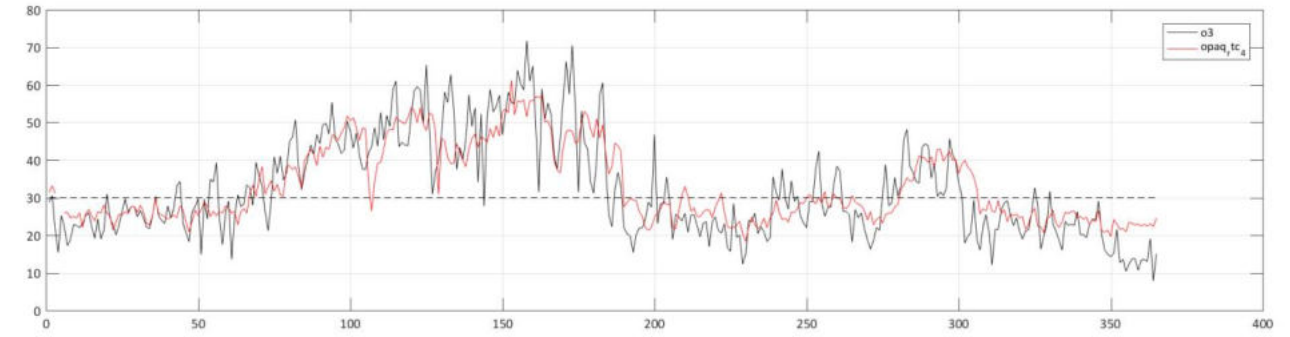

o3_dayavg_Delhi_opaq_rtc_1
Scatter mod vs. obs
RMSE = 7.10
BIAS = 0.45
R² = 0.71
opaq_rtc_1
o3
Target plot (v1)
BIAS/σ_O
CRMSE/σ_O
R
SD
QQ plot
Mod. quantiles
Obs. quantiles
o3
opaq_rtc_1


o3_dayavg_Delhi_opaq_rtc_2
Scatter mod vs. obs
RMSE = 7.39
BIAS = 1.18
R² = 0.69
opaq_rtc_2
o3
Target plot (v1)
BIAS/σ_O
CRMSE/σ_O
R
SD
QQ plot
Mod. quantiles
Obs. quantiles
o3
opaq_rtc_2

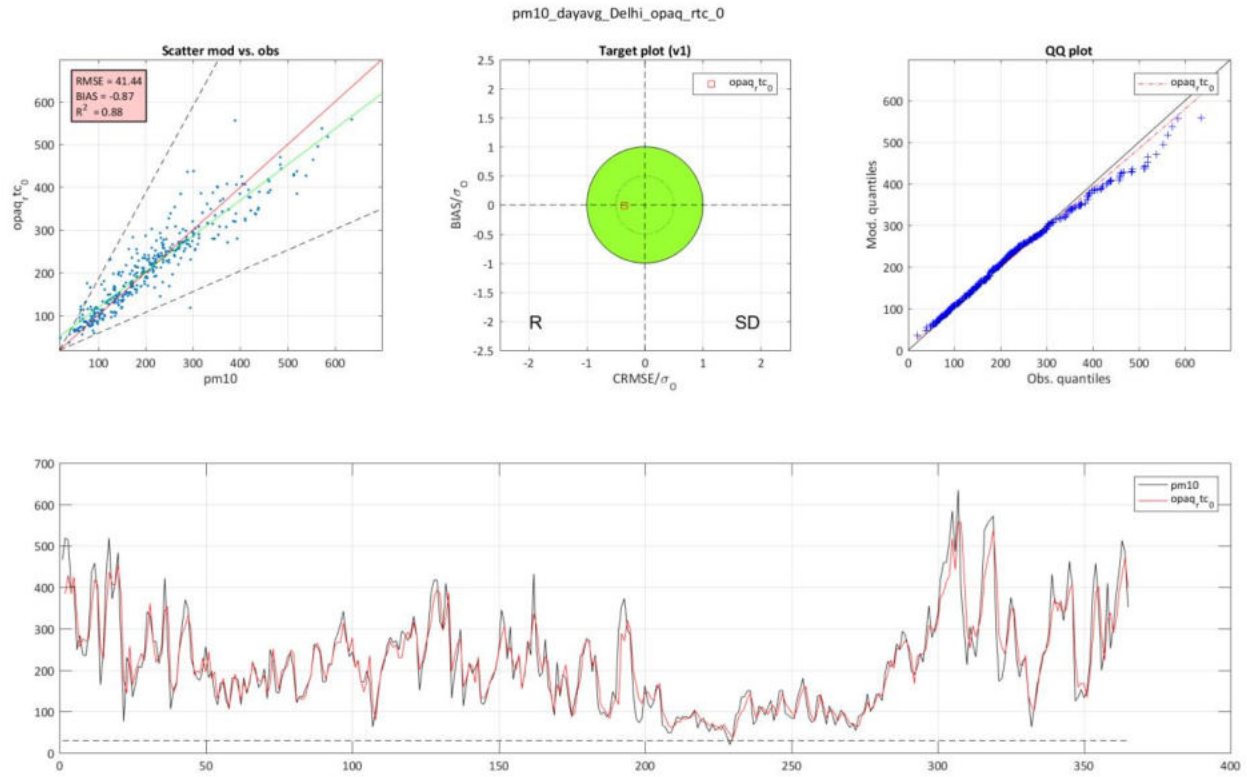

o3_dayavg_Delhi_opaq_rtc_3
Scatter mod vs. obs
RMSE = 7.38
BIAS = 0.13
R² = 0.69
opaq_rtc_3
o3
Target plot (v1)
opaq_rtc_3
BIAS/σ_O
CRMSE/σ_O
R
SD
QQ plot
opaq_rtc_3
Mod. quantiles
Obs. quantiles

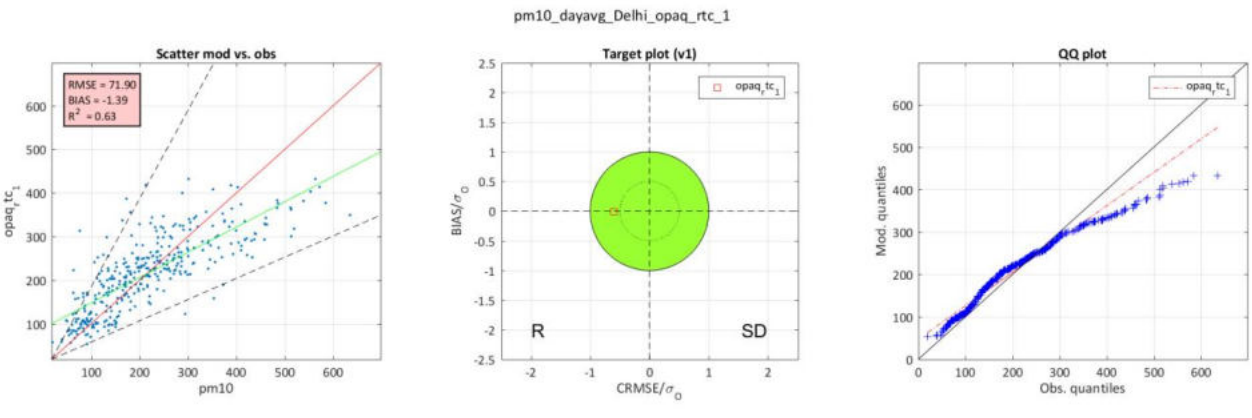

o3
opaq_rtc_3

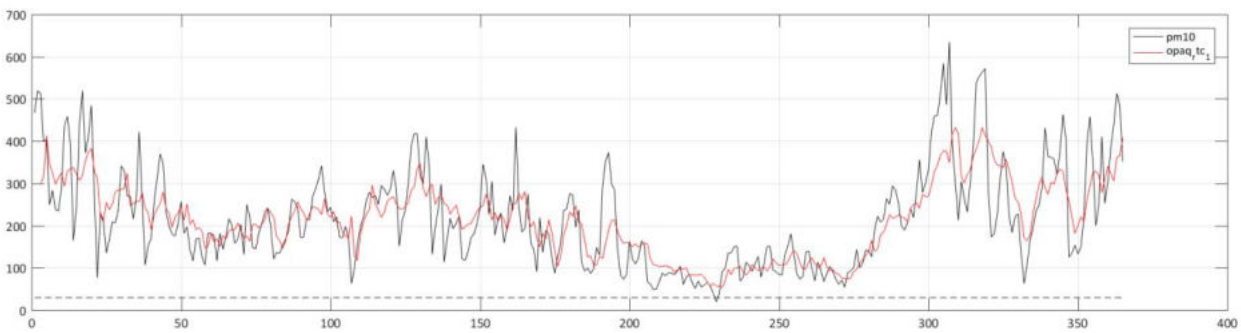

o3_dayavg_Delhi_opaq_rtc_4
Scatter mod vs. obs
RMSE = 7.51
BIAS = 0.70
R² = 0.68
opaq_rtc_4
o3
Target plot (v1)
opaq_rtc_4
BIAS/σ_O
CRMSE/σ_O
R
SD
QQ plot
opaq_rtc_4
Mod. quantiles
Obs. quantiles

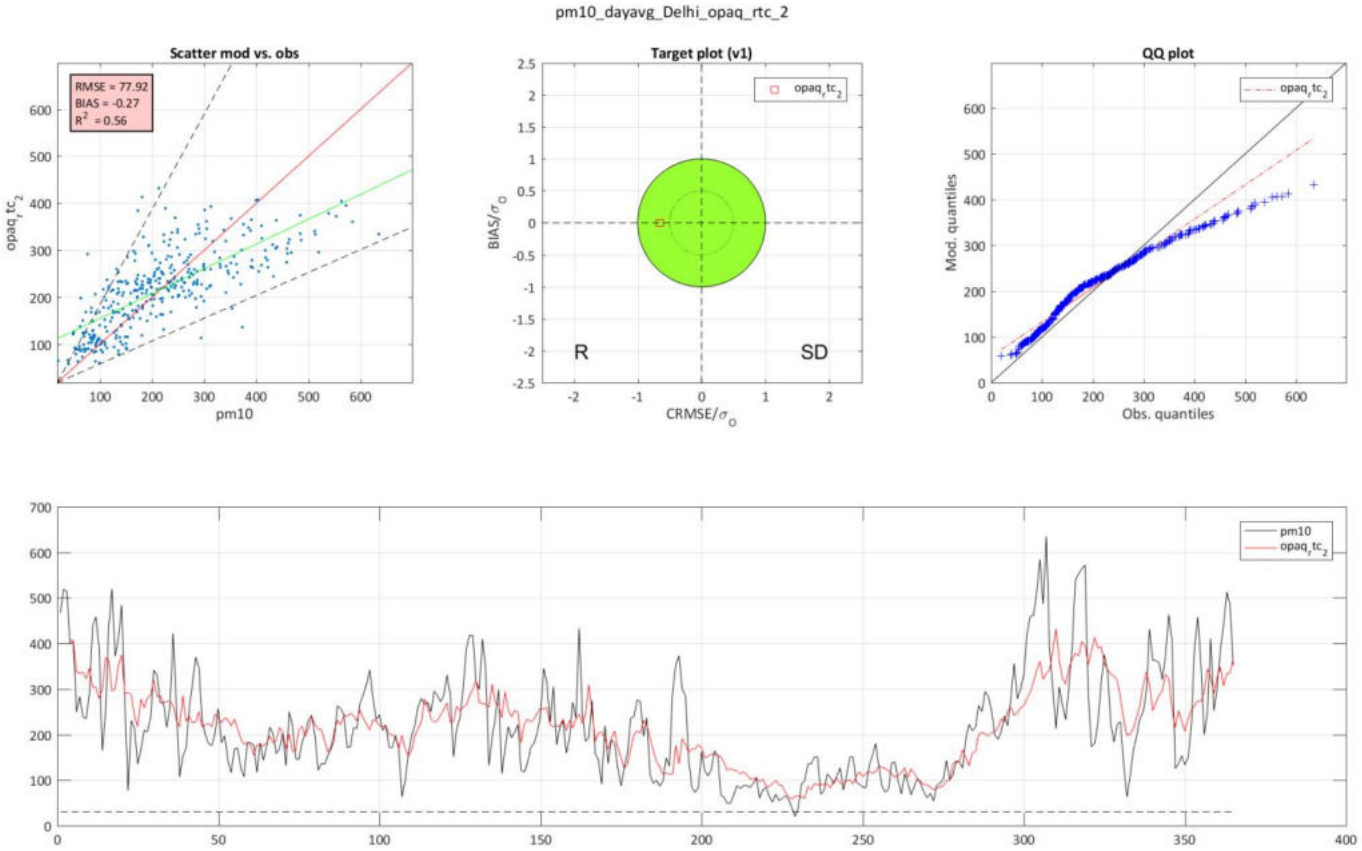

o3
opaq_rtc_4

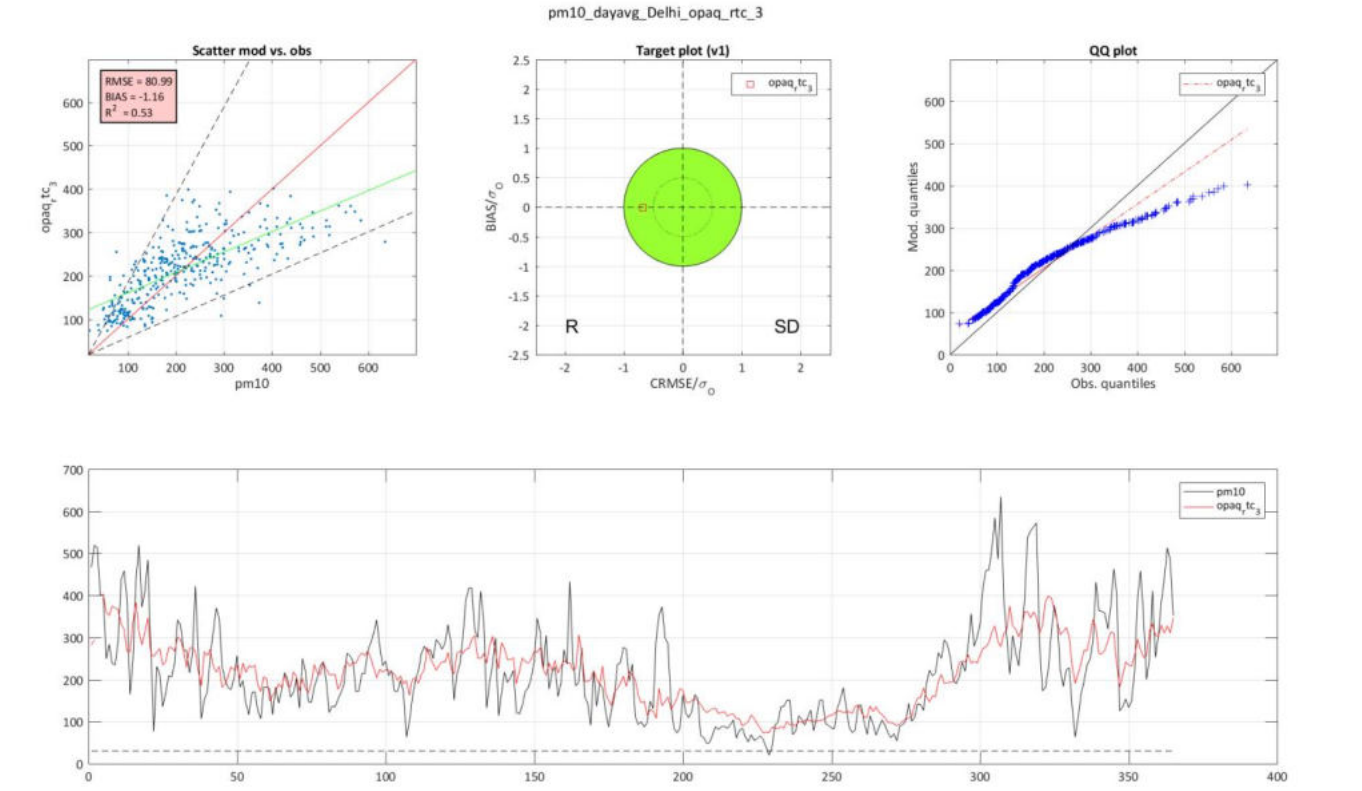
PM_{10} (Arranged from Day0 to Day4)

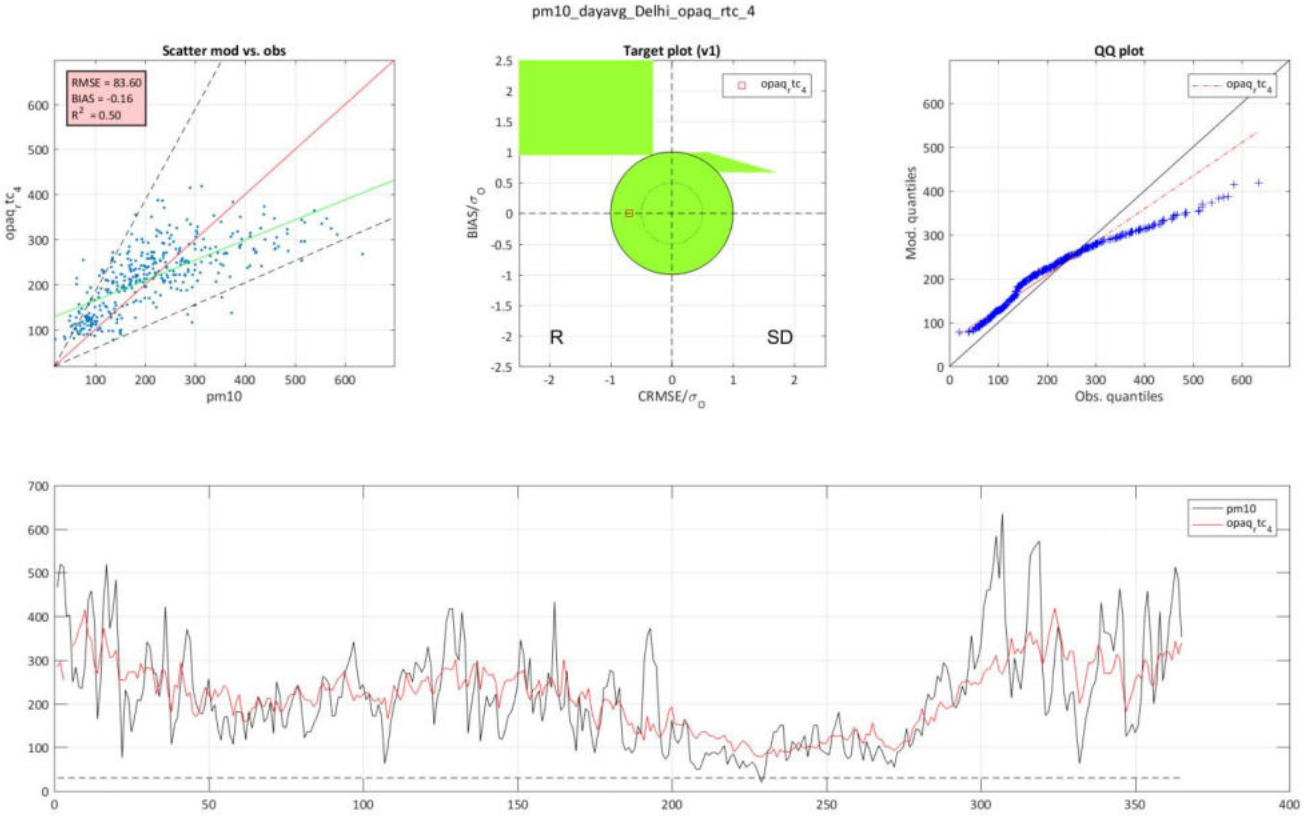




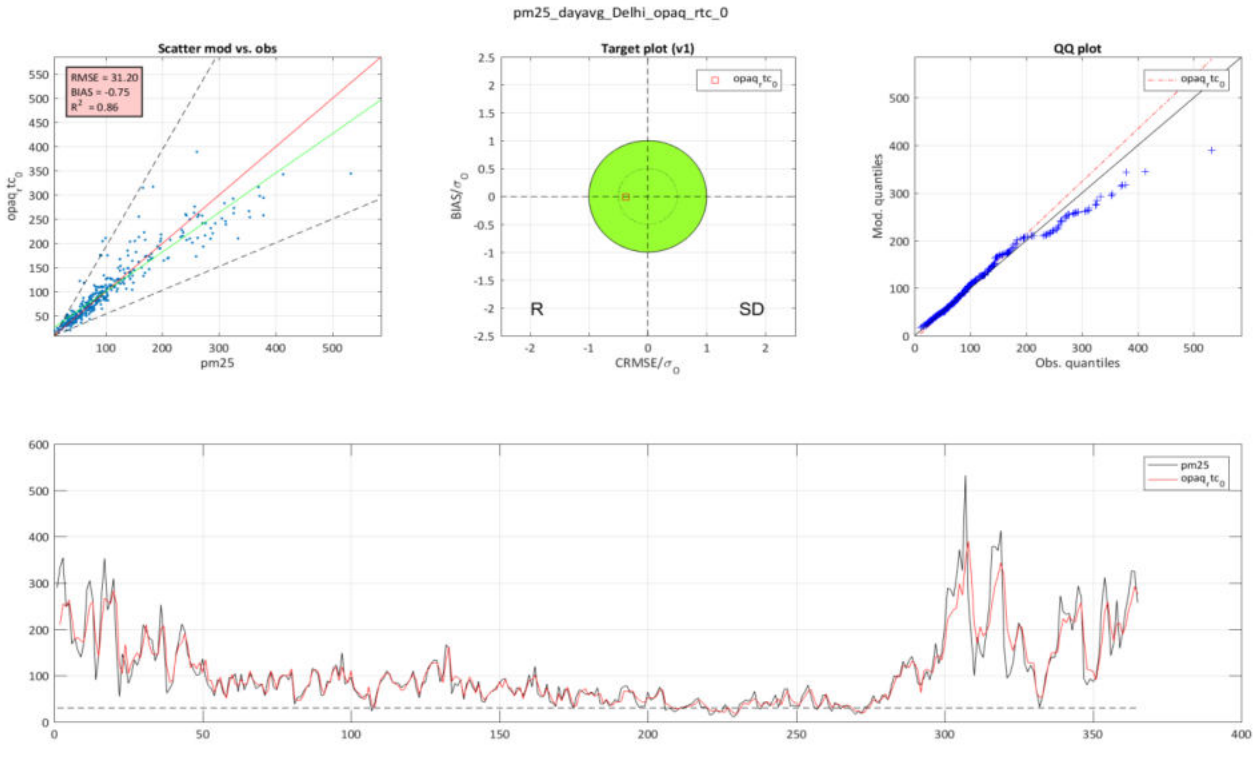



pm10_dayavg_Delhi_opaq_rtc_2
Scatter mod vs. obs
RMSE = 77.92
BIAS = -0.27
R² = 0.56
pm10
Target plot (v1)
R
SD
CRMSE/σ_O
QQ plot
Mod. quantiles
Obs. quantiles
pm10

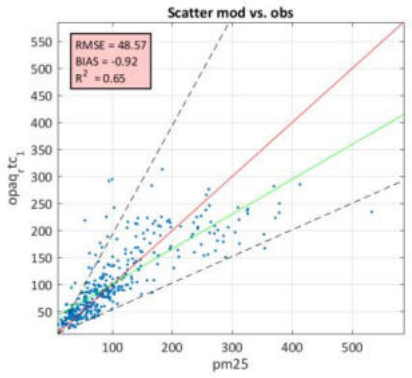

pm10_dayavg_Delhi_opaq_rtc_3
Scatter mod vs. obs
RMSE = 80.99
BIAS = -1.16
R² = 0.53
pm10
Target plot (v1)
R
SD
CRMSE/σ_O
QQ plot
Mod. quantiles
Obs. quantiles
pm10

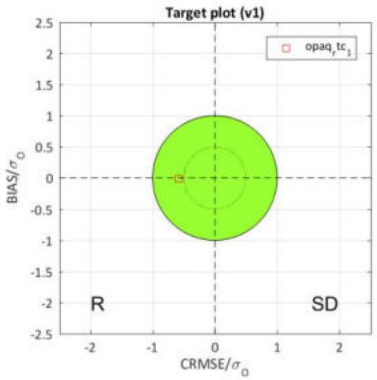


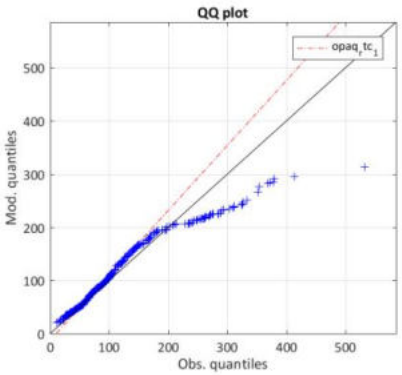
$PM_{2.5}$ (Arranged from Day0 to Day4)

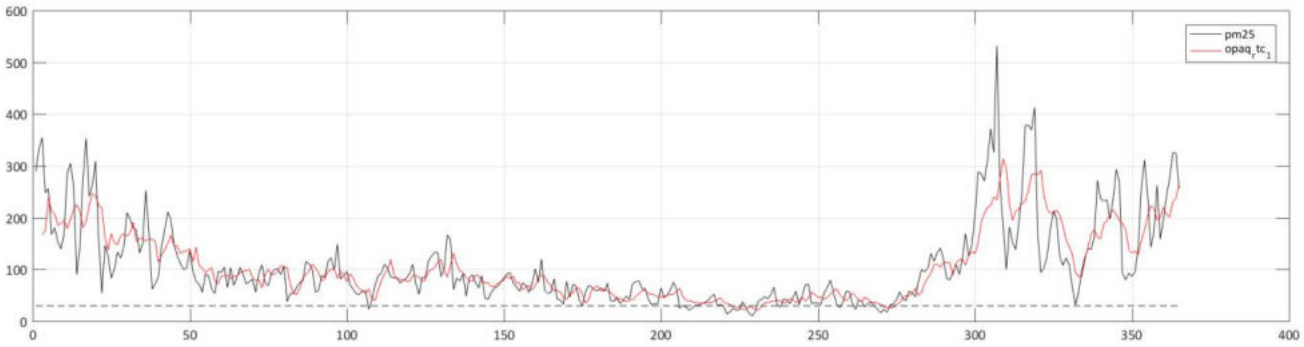


pm25_dayavg_Delhi_opaq_rtc_1

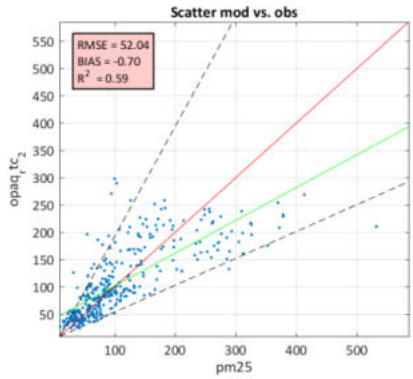


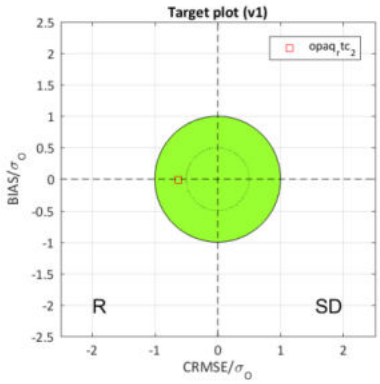


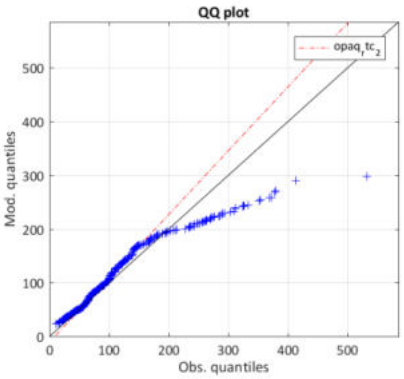


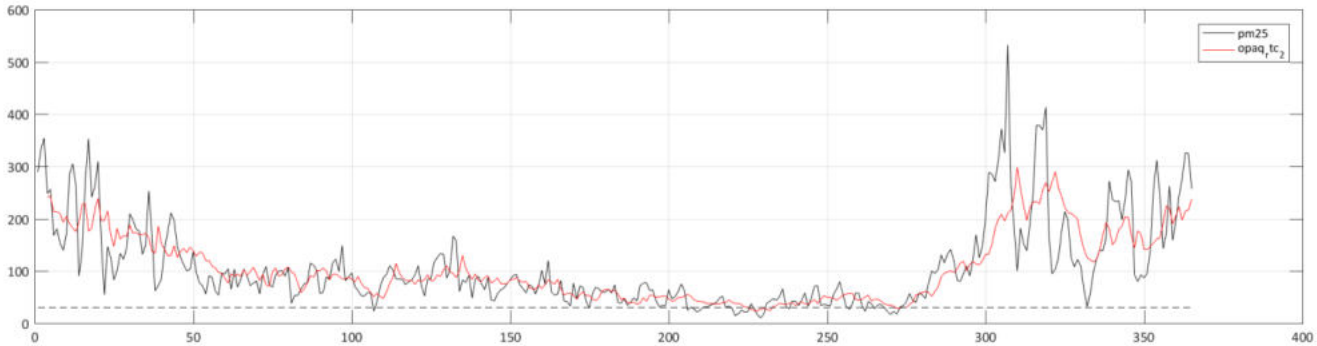


pm25_dayavg_Delhi_opaq_rtc_2

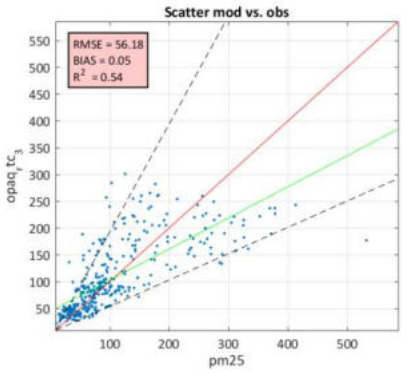


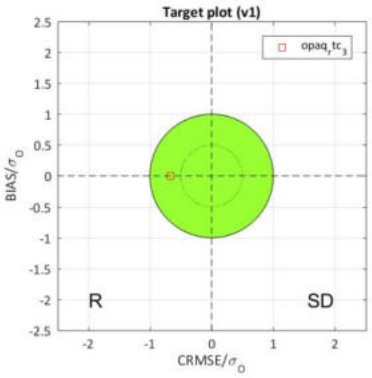


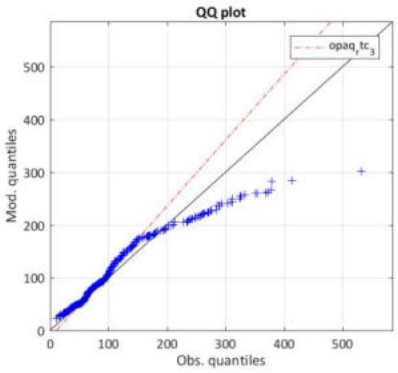


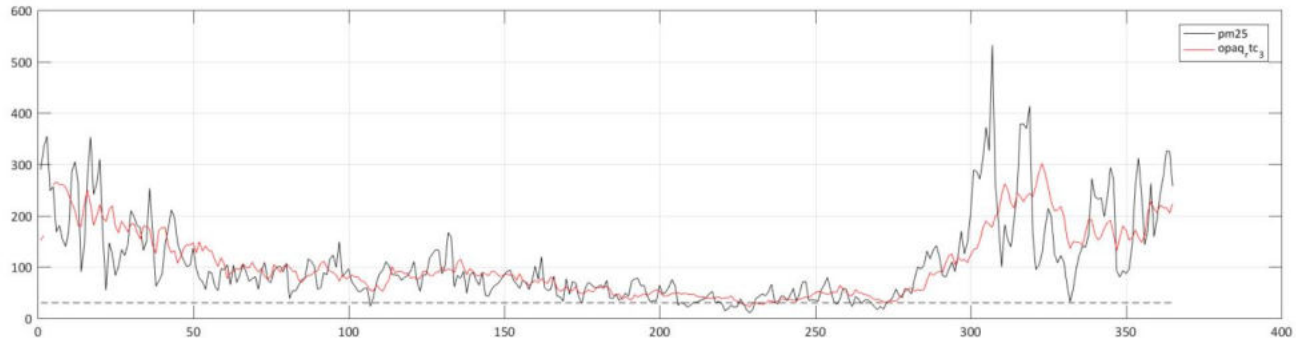


pm25_dayavg_Delhi_opaq_rtc_3

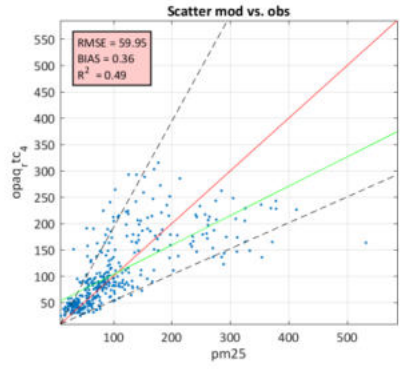


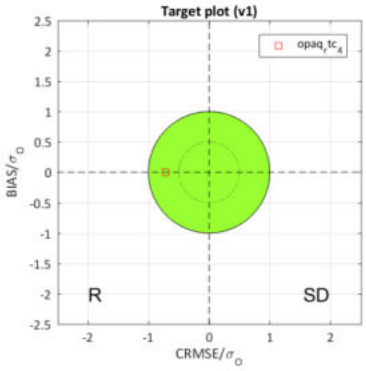


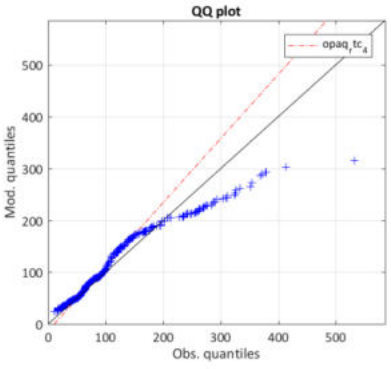


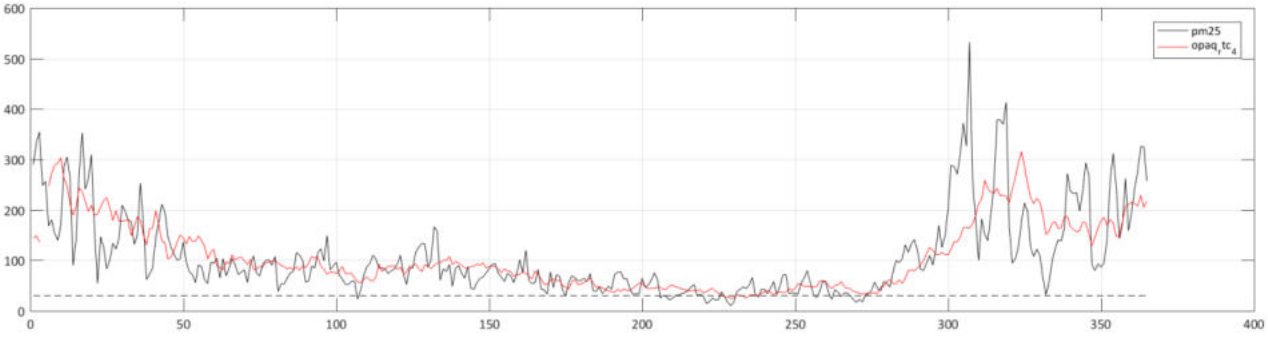


pm25_dayavg_Delhi_opaq_rtc_4









About TERI

A unique developing country institution, TERI is deeply committed to every aspect of sustainable development. From providing environment-friendly solutions to rural energy problems to helping shape the development of the Indian oil and gas sector; from tackling global climate change issues across many continents to enhancing forest conservation efforts among local communities; from advancing solutions to growing urban transport and air pollution problems to promoting energy efficiency in the Indian industry, the emphasis has always been on finding innovative solutions to make the world a better place to live in. However, while TERI's vision is global, its roots are firmly entrenched in Indian soil. All activities in TERI move from formulating local- and national-level strategies to suggesting global solutions to critical energy and environment-related issues. TERI has grown to establish a presence in not only different corners and regions of India, but is perhaps the only developing country institution to have established a presence in North America and Europe and on the Asian continent in Japan, Malaysia, and the Gulf.

TERI possesses rich and varied experience in the electricity/energy sector in India and abroad, and has been providing assistance on a range of activities to public, private, and international clients. It offers invaluable expertise in the fields of power, coal and hydrocarbons and has extensive experience on regulatory and tariff issues, policy and institutional issues. TERI has been at the forefront in providing expertise and professional services to national and international clients. TERI has been closely working with utilities, regulatory commissions, government, bilateral and multilateral organizations (The World Bank, ADB, JBIC, DFID, and USAID, among many others) in the past. This has been possible since TERI has multidisciplinary expertise comprising of economist, technical, social, environmental, and management.

The Energy and Resources Institute
www. teriin.org