

# Prediction and assessment of change in PM2.5 during COVID-19 lockdown using remote sensing and deep learning approach: A case study of Kanpur city

Santanu Mallik (✉ [coolshan02@gmail.com](mailto:coolshan02@gmail.com))

National Institute of Technology Agartala

**Shivam Soni**

Central Water Commission, Jammu

**Krishanu Podder**

CDAC (PGDAC), Pune

**Umesh Mishra**

National Institute of Technology Agartala

**Maqsood Ahamed**

King Abdullah Institute for Nanotechnology, King Saud University

---

## Research Article

**Keywords:** Aerosol optical depth, Air pollution, MODIS, Multi-linear regression, Artificial neural network.

**Posted Date:** October 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-88484/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

In the last few months, the corona virus crises have changed the world completely. In India, looking at the horrific condition and rate of increase of COVID-19 positive cases, the central government has imposed strict lockdown from dated March 24 throughout the country. At the same time, with citizens quarantined and grinding halt in industrial activities has led to some inadvertent benefits and change in air pollution level is one of them. The present study is undertaken with an aim to predict and assess the change in PM<sub>2.5</sub> level of the Kanpur city using MODIS Aerosol Optical Depth (AOD) data and meteorological parameters during various stages of COVID-19 lockdown. Subsequently, four different methods such as linear regression (LR), multilinear regression (MLR), artificial neural network (ANN), and a hybrid method were used and compared for the prediction of PM<sub>2.5</sub>. The best prediction method was further used for the spatial mapping of PM<sub>2.5</sub> at different stages of lockdown (pre-lockdown, during lockdown and post lockdown). The result of cross-validation shows that the hybrid method i.e. (MLR+ANN) outperforms all other methods and gives the highest coefficient of determination  $R^2$  value of 0.961 followed by ANN, MLR, and LR with 0.895, 0.246, 0.016. The spatial mapping of PM<sub>2.5</sub> during the lockdown shows a significant reduction of 26% in PM<sub>2.5</sub> concentration in comparison to pre lockdown. The methodology developed from the current study can also be used in other regions for predicting and analyzing the spatio-temporal variation of PM<sub>2.5</sub>.

## 1. Introduction

The outbreak of the novel corona virus has come into the limelight on December 31<sup>st</sup> 2019 when China has reported a few bundles of cases on pneumonia in Wuhan City in Hubei Province. The deadly virus then rapidly spread across all the corners of the world and soon it was declared as a pandemic by the WHO. According to the WHO report of 23 March, more than 16,000 people have died worldwide, and more than 375,000 have been infected (The Hindu)(World Health Organization 2020). Looking at the rapid spread of COVID-19 on March 24, the Indian government has declared lockdown and canceled all the foreign and domestic flights. Since then a chaos and frenzy free environment is observed across all the major cities of India. Several environmental factors such as noise, water, air pollution show improvement in quality due to this lockdown. In cities, most of the air pollution is due to vehicular emissions and amidst the devastating COVID-19 lockdown there had been a huge reduction in number of vehicles which resulted in a reduction of air pollution (Dantas et al. 2020; Lokhandwala and Gautam 2020; Mahato and Ghosh 2020; Rodríguez-Urrego and Rodríguez-Urrego 2020).

PM<sub>2.5</sub> is one of the most critical air pollutants, which enters the heart and stays there for a long time resulting in severe cardiovascular and respiratory diseases (Chu et al. 2016). Various state and central monitoring agencies carry out the continuous ambient PM<sub>2.5</sub> monitoring throughout India. The retrieval of PM<sub>2.5</sub> from the ground based station is having certain constrain due to its limited spatial converge , high cost involvement, and continuous calibration of these device is a quite tedious (Chelani 2019; Zha et al. 2010). So, the usage of the remote sensing based method using Aerosol Optical Depth (AOD) values for PM<sub>2.5</sub> retrieval has shown a great potential to overcome these problems. The AOD measures the

amount of light transmittance through a unit section of an atmospheric column (Ahmad et al. 2019; Zhao et al. 2018).

Several statistical and mathematical methods were used to estimate the PM<sub>2.5</sub> from AOD data, out of which the linear regression was the mostly used method. Somehow due to limitation and accuracy problem this method is not in use in the recent literatures (Wang and Christopher 2003; Zha et al. 2010). Among all the other methods, the recent studies are using the multiple linear Regression (MLR), Chemical Transport Model (CTM), and Mixed-Effect Model (MEM) method using various secondary data and carrying out simulation for the better prediction of PM<sub>2.5</sub>.

A handful of researchers have reported significant improvement in the prediction of ground-level PM<sub>2.5</sub> by using meteorological data along with AOD (Chelani 2019; Zha et al. 2010; Zhao et al. 2018). A linear mixed effect model along with meteorological parameters and NO<sub>2</sub> column density measurements was incorporated by (Zheng et al. 2016) to improve the prediction value. (Van Donkelaar et al., 2006) used a global chemical transport model (GEOS-CHEM) and simulate the factors effecting AOD and PM<sub>2.5</sub> for better prediction of PM<sub>2.5</sub>. (Murray et al. 2019) has estimated PM<sub>2.5</sub> by combining both the AOD values and CTM simulation using statistical method and results shows improvement in R<sup>2</sup> value. (Sorek-Hamer et al. 2015) used the AOD product, mixed-effect model, and the daily calibration approach to predict PM<sub>2.5</sub> and observed a considerable improvement in prediction of PM<sub>2.5</sub> for high reflectance regions.

The growing importance of the machine and deep learning methods over linear methods to predict PM<sub>2.5</sub> using satellite and other metrological parameters is gaining more attention due to its ability to handle large amounts of data. The recent studies are using various deep learning and machine learning-based method such as geographically Weighted Regression (GWR), convolution neural network (CNN), artificial neural network (ANN), gradient boosting machine (GBM), random forest (RF), k-nearest neighbor (KNN), decision Tree regressor CART approach and so on so forth. (Hu et al. 2013; Song et al. 2014; You et al. 2016) used a geographically weighted regression (GWR) model to estimate the PM<sub>2.5</sub> concentration based on AOD values and meteorological parameters. The study conducted by (Ahmad et al. 2019) use both MLR and ANN approach by taking the support of land use data and meteorological parameters and obtained an enhanced performance by adding the same. (Doreswamy et al. 2020) used various machine learning methods such as random forest regressor, gradient boosting regressor, k neighbors regressor, MLP regressor, and decision tree regressor CART and shown a better prediction result as compare to the traditional method. (Joharestani et al. 2019) has implemented random forest, extreme gradient boosting, and machine learning (ML) approaches along with 23 features and obtained better prediction results. (Yazdi et al. 2020) has used and compared various machine learning approach and found out that the random forest (RF) technique performs the best in comparison with all other method used. (Shahriar et al. 2020) has used the linear-support vector machine (L-SVM), medium Gaussian-support vector machine (M-SVM), Gaussian process regression (GPR), artificial neural network (ANN), random forest regression (RFR), and a time series model namely PROPHE and found out that GPR and ANN perform better in (Ventura et al. 2019) has applied two models, Holt–Winters (HW) and artificial neural network (ANN), using PM 2.5 concentration time series and found ANN performs better with less RMSE error.

Major shortcomings associated with all these prediction methods applied in the above studies is that, only graphical representation are reported in these literatures along with cross validation results. Although, PM2.5 is a spatial data and it changes with respect to space and time and most of study does not show the spatial representation of the PM2.5. A spatial representation of PM2.5 along with graphical representation may enhance the work and it will help the readers to better understand the spatial distribution of PM2.5 across the study area. So, to solve this quest this study uses a simple and easily understandable hybrid method which not only predict the PM2.5 but also produce the spatial maps for better visualization and representation of spatial distribution of PM2.5. Furthermore, the study area, Kanpur's have poor air quality index in comparison with major cities in India and limited literatures are available from this study area. The methodology developed in this study will aid in the monitoring the spatial variation of PM2.5. In addition, the impact of various anthropogenic activities and natural phenomena on the concentration of PM2.5 can also be monitored.

The present study is initiated by setting these objectives (i) to develop statistical, deep learning, and hybrid models for prediction of PM2.5 (ii) to compare the performance of LR, MLR, ANN, and Hybrid method i.e. (MLR+ANN) models and find the best model for prediction (iii) to analyze the spatial distribution of PM2.5 during various phases of lockdown for COVID-19 in and around Kanpur city.

## 2. Study Area

Kanpur is the industrial capital of Uttar Pradesh and a major town in the northern state of India. This town is situated on the south bank of river Ganga, located 80 km west of Lucknow. Kanpur Fig 1 is having an total area of around 10,863 Km<sup>2</sup> and a population of 45,81,268. (<https://kanpurnagar.nic.in/>). The districts of Kanpur are divided into two parts Kanpur urban and Kanpur Rural. The average low and high temperature of Kanpur varies from around 8.2 to 39.8<sup>0</sup> C respectively (Indian Meteorological Department).

## 3. Materials And Methods

### 3.1 Data Preparation

Ground-based measured PM2.5 concentration, satellite-derived MODIS AODs data, and meteorological data (wind speed, temperature, rainfall, relative humidity, and mixing height) were used in this study. Table 1 shows the details of data and its source used in this study. The selection of meteorological data was based on expert advice and literature review. Exploratory data analysis (EDA) was carried out to understand the data characteristics. Fig 2 shows the overall methodology and graphical representation of work used in this study.

**Table 1** The list of data used in this study

Variables	Units	Data Source
PM2.5	$\mu\text{g}/\text{m}^3$	<a href="https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing">https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing</a>
AOD	nm	<a href="https://ladsweb.modaps.eosdis.nasa.gov/">https://ladsweb.modaps.eosdis.nasa.gov/</a>
Temperature	$^{\circ}\text{C}$	<a href="https://mausam.imd.gov.in/">https://mausam.imd.gov.in/</a>
Rainfall	mm	<a href="https://mausam.imd.gov.in/">https://mausam.imd.gov.in/</a>
Relative humidity	$\text{gm}/\text{m}^3$	<a href="https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing">https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing</a>
Wind speed	m/s	<a href="https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing">https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing</a>
Mixing height	m	<a href="https://www.ready.noaa.gov/HYSPLIT.php">https://www.ready.noaa.gov/HYSPLIT.php</a>

### 3.1.1 PM2.5 field data

Ground station data of daily average PM2.5 were collected for the month of February 1<sup>st</sup> (pre-lockdown) to August 10<sup>th</sup> (during Post lockdown) 2020 from central pollution control board (CPCB). A total 163 number of data were collected for the Kanpur city. Nehru Park Kanpur (Latitude = 26.4703136 & Longitude = 80.3119863) data was used in this study.

### 3.1.2 Aerosol optical depth (AOD)

The MODIS sensors with spatial resolution and temporal of 250-1km and 1-2 days aboard on the Terra and Aqua satellites provide the aerosol optical depth (AOD) at 550nm. The storage format of MODIS\_L2 data is .hdf, and looking at the heterogeneity of the study area, the combined Dark Target (DT) algorithm and Deep blue over land and ocean was used for the retrievals of aerosol data with spatial resolution of 3 and 10 km. The exact coordinate of the study area was used to extract the AOD value. The data obtained was screen out to remove invalid MODIS files. More details about MODIS satellite data retrieval are available on the article by (Kloog et al., 2012). The 10 km data shows a strong correlation with PM2.5 data so this data was further used for analysis.

### 3.1.3 Meteorological data

Meteorological data sets (wind speed, temperature, rainfall, and relative humidity) were acquired from the Indian meteorological department and CPCB. The mixing height data was calculated using NOAA (National Oceanography and Atmospheric Agency) based HYSPLIT (HYbrid Single-particle Lagrangian Integrated Trajectory) model (Draxler, R. R., & Rolph 2010). Temperature enhances the production of PM2.5 concentration by photochemical reactions and relative humidity directly impacts on the scattering of the particle (due to hygroscopic nature). Similarly, wind speed and mixing height affect the pollutant mixing and distribution of aerosols (Gupta et al., 2006; Soni et al., 2018).

## 3.2 Prediction models

### 3.2.1 Linear regressions (LR)

The simple Linear regression (LR) method was used to model spatially varying relationships between the dependent (PM2.5) and explanatory variable (AOD), this relationship is acquired by fitting a linear curve to the data. The LR method works on the assumption of spatial stationary.

$$\mathbf{PM2.5} = \boldsymbol{\mu} * \mathbf{AOD} + \boldsymbol{\lambda} \quad (1)$$

Where, **PM2.5** measure concentration in the ground ( $\mu\text{g}/\text{m}^3$ ),  $\boldsymbol{\mu}$  and  $\boldsymbol{\lambda}$  are the slope and intercept.

Minitab2017 software was used for performing the LR model. The use of single predictor in LR and the assumption made causes a significant error in prediction value (Ahmad et al. 2019). Therefore, the multiple linear regressions (MLR) method was used with more number of explanatory variables.

### 3.2.2 Multiple linear regressions (MLR)

The Multiple linear regressions (MLR) is a statistical regression approach, which uses multiple depended and explanatory variable for finding a linear relationship between them.

The MLR equation can be expressed as follows

$$\mathbf{Y} = \mathbf{a}_1\mathbf{X}_1 + \mathbf{a}_2\mathbf{X}_2 + \mathbf{a}_3\mathbf{X}_3 + \dots + \mathbf{a}_n\mathbf{X}_n + \mathbf{C} \quad (2)$$

Where, Y is dependent variable (PM2.5),  $X_1, X_2, \dots, X_n$  are explanatory variables (i.e. meteorological variables) and  $a_1, a_2, \dots, a_n$  is the regression co-efficient that indication variable contribution and C is intercept.

To optimize the performance of MLR the most important factor affecting the regression prediction and multi-co linearity was checked using stepwise forward variable selection according to the lowest Akaike information criterion (AIC) and the variance inflation factor (VIF) (Knoll et al. 2019). Minitab2017 software was used for carry out the MLR model.

### 3.2.3 Artificial neural network (ANN)

Artificial neural network (ANN) is a non-linear supervised deep learning method, which tends to mimic the human brain and biological nervous system (Amanollahi and Ausati, 2020; Ordieres et al., 2005; Yao et al., 2012). The capability of learning a complex relationship problem between input and output parameters makes the ANN model more flexible and more usable in many multidisciplinary fields (Emamgholizadeh et al. 2017; Mallik et al. 2020).

The most popular Multilayer perceptron (MLP) network with a feed-forward back propagation algorithm was used in this study (Ahmad et al., 2019; Amanollahi and Ausati, 2020; Fenget al., 2015; Ordieres et al., 2005; Yao et al., 2012). Prior to the selection of the network, the data were normalized to increase the processing speed and accuracy (Feng et al., 2015). Leungberg-Marquard's learning function was considered while training the ANN model. Meteorological data and AOD data are used as input variables for the input layer. Numerous trials and errors (network parameters, variables) were carried out to get the best network. The number of hidden layers and neurons was ranged from 2-3 and 5 to 25 with a step size of one. (MATLAB R2013b) the software was used for performing ANN models.

### 3.2.4 Hybrid- Method (MLR+ANN)

A hybrid method was used in this study by combining the MLR and ANN method. Fig 3 depicts the steps involved in this method. The residuals obtained from the MLR method were modeled using the ANN method. The advantage of the hybrid method is it has the strength of both the linear and non-linear methods. This method not only helps in removing the correlation present in residual but also leads to better predictions and lower error (Chelani 2019).

### 3.3 Model evaluation and validation

The overall accuracy of the all the models were evaluated by coefficient of determination ( $R^2$ ) and the best network geometry of the ANN and hybrid (MLR+ANN) method was selected on least mean square error (MSE) value

$$MSE = \frac{\sum_{i=1}^N (PM2.5_m - PM2.5_p)^2}{N} \quad (3)$$

Where,  $PM2.5_m$  and  $PM2.5_p$  is the measured and predicted value of  $PM2.5$  and  $N$  is the number of data point

### 3.4 Spatial mapping of $PM2.5$

The best model obtained for the prediction of  $PM2.5$  was used to map the  $PM2.5$  concentration throughout the various stages of lockdown (Pre and during lockdown). The Python 3.7 (ipython) console of spyder IDE and packages such as pyhdf, numpy, sys and, mpl\_toolkits.basemap was used for the loading, analyzing and producing the spatial mapping of  $PM2.5$ .

## 4. Result And Discussion

### 4.1 Descriptive statistics

The descriptive statistics of the data used are shown in Table 2. It is observed that except rainfall and mixing height all other variables are positively skewed. The Kurtosis values show that only  $PM2.5$  and

AOD values are positive which indicates the profusion of an outlier in the data. The PM2.5 value in the study area ranges between 20.69 and 128.02 $\mu\text{g}/\text{m}^3$  with a mean value of 50.32 $\mu\text{g}/\text{m}^3$ . Whereas, the mixing height and the AOD show the maximum and minimum deviation from the mean.

A Correlation is a statistical technique that shows whether and how the variables are linearly associated with each other. Fig 4 shows the Spearman correlation plots between different parameters. The Spearman correlation coefficient was used over the commonly used Pearson correlation factor because it does constrain with normally distributed parameters. The diagonal square boxes in the figure are shown in yellow color indicates correlation with itself. Whereas dark color one indicated a higher correlation between the variables. The highest positive and negative correlation was observed between PM2.5 with AOD and wind speed.

**Table 2.** Descriptive statistics

	AOD	PM2.5	Temperature	Rainfall	Relative humidity	Wind speed	Mixing height
<b>Min</b>	0.072	20.69	33.69	19.33	40.38	0.5	1682
<b>Max</b>	1.66	128.02	47.19	36.8	87.56	3.71	5046
<b>Mean</b>	0.545	50.32	37.41	28.33	62.457	1.841	3504.02
<b>SD</b>	0.329	18.73	3.66	4.294	10.304	0.886	799.19
<b>Skewness</b>	1.07	1.22	1.103	-0.053	0.246	0.324	-0.158
<b>Kurtosis</b>	1.508	3.021	-0.059	-0.605	-0.763	-0.789	-0.699

## 4.2 Estimating PM2.5

Various methods were used in this study to estimate the ground-based PM2.5 using MODIS AOD data and meteorological data. The coefficient of determination ( $R^2$ ) was used as a marker to determine, whether the general models contribute to the overall variance in data. As a traditional approach, statistical based LR was used to predict PM2.5 using the AOD values but the result of the correlation shows a very poor value of  $R^2$  i.e. 0.016. Furthermore, to improve the prediction the MLR analysis was used and the  $R^2$  score was considerably increased to 0.246. Out of six, only five variable i.e. AOD, temperature, Relative humidity, Wind speed, and Mixing height were used to obtained the  $R^2$  value. The variable rainfall was omitted because of moderate multi co-linearity and the highest probability (P) value. Table 1S(a,b) shows the variance inflation factor and P- values for the MLR model used as well as the residual vs. fitted plot for MLR is also shown in Fig 1S.

Unlike a purely statistical method, a deep learning approach was employed in this study, which takes multiple parameters into account in a single model. Artificial Neural Networks (ANN) is one of the most popular classifier methods to predict PM2.5 from meteorological data. Fig 2S shows the best ANN structure selected based on the least MSE value. In comparison with the MLR model, the use of the ANN model causes a significant rise in the  $R^2$  value from 0.246 to 0.895. In order to achieve better prediction accuracy a hybrid model was developed using a combination of MLR and ANN method, which has resulted in increase of  $R^2$  value by 7.37% in comparison to the ANN model. Table 3, Fig 5 shows the

comparison between all the methods used in this study for the prediction of PM2.5. Thus the results obtained from comparing these method shows that the hybrid method outperform all other method with highest R<sup>2</sup> value of 0.961. Similar result was also reported by (Doreswamy et al. 2020) using the Gradient Boosting regression method. As all the methods used are data driven model may be inclusion of more number of data may have provided better accuracy.

**Table 3.** Comparison between all the models used for prediction of PM2.5

<b>Name of the model</b>	<b>Parameters used</b>	<b>R<sup>2</sup> Value</b>
<b>Linear Regression (LR)</b>	AOD	0.016
<b>Multiple Linear Regression (MLR)</b>	AOD,T,R,RH,WS,MH	0.221
	AOD,T, RH,WS,MH	0.246
<b>Artificial Neural Network (ANN)</b>	AOD,T,R,RH,WS,MH	0.895
<b>Hybrid (MLR+ANN)</b>	AOD,T,R,RH,WS,MH	<b>0.961</b>

Where, AOD- Aerosol optical depth , T- Temperature, RH- Relative Humidity, WS - Wind Speed, MH- Mixing Height

Even though the results obtained by using this hybrid method for prediction of PM2.5 is not too high as 0.97 obtained by (Amanollahi and Ausati 2020) still, comparatively better than some of the results obtained by (Joharestani et al. 2019; Yazdi et al. 2020). Thus, this method can easily be implemented because of its robustness as it has can be used for prediction as well as spatial representation of PM2.5 compared to other approaches used.

The variability of the predicted PM2.5 using various models with respect to actual observed PM2.5 is shown using a Box plot in Fig 6. The figure indicated that the center, spread value of the observed value of PM2.5 is almost equal to that of the Hybrid method. The Maximum value is almost the same for MLR and hybrid method whereas the minimum value was slightly underestimated.

#### 4.2 Spatial mapping of PM2.5 during (Pre-lockdown, lockdown)

In this section, we have shown the changes in the concentration of PM2.5 during pre (March), lockdown (April) and post lockdown (July-August). The number of data collected during the lockdown and pre, lockdown and, post lockdown is not the same so, an equal number of data was used for the better visualization of changes that occurred in the different range of PM2.5.

Fig 7 shows the frequency distribution of PM2.5 concentration of the Kanpur city during different phases of the lockdown. It can be seen from the figure that the frequency of higher concentration of PM2.5 during lockdown has significantly reduced by almost 43%, whereas the frequency of lower concentration

of PM2.5 has increased by almost 42% including all the ranges from minimal impact to minor problem. During post lockdown all the data are within lower concentration of PM2.5.

The comparison of average change in concentration of PM2.5 for the whole data set collected from March to August has revealed that PM2.5 has decreased by approximately 26% during the lockdown (March – May), and there is further reduction in PM2.5 during the post lockdown by 34% (June-August). Similar result is also reported by (Agarwal et al. 2020) on nearby study area which supports our findings. Furthermore, for more clarification whether the change in PM2.5 is due to season variation the last year 2019 PM2.5 data were obtained for the same period and the changes in level of PM2.5 was also monitored shown in Fig.8. It can be seen from the observation that last year during March to May there is only 4% reduction in average PM2.5 concentration, whereas, the further reduction in PM2.5 during June to August is 47%. As compare to current year the reduction in PM2.5 previous year during June to August is considerably high, this is because of increase private 2 or 4 wheeler vehicles. The increase in usage of the private vehicle may be due to restriction imposed in the movement of public transport and people are more aware with the fact of getting contaminated from the virus.

Fig 9(a,b,c) shows the spatial mapping to understand the spatial variability of PM2.5 along with change in PM2.5 level due to the lockdown in an around Kanpur city. A total nine number of maps were prepared three for each pre lockdown, during lockdown, and post lockdown stage. The maps show that during the pre lockdown stages the PM2.5 is mostly under unhealthy to very unhealthy zone, whereas the impact of lockdown due to COVID-19 pandemic has drastically altered the PM2.5 concentration from moderate to good. This is mostly due to the total absence of all emission-based industries and no vehicles on the road during different phases during lockdown. Whereas, in the Post lockdown period the PM2.5 level is in the moderate zone. The spatial mapping helpss us in undrstanding the spatial distribution of PM2.5 in and around any study area. This methodology developed can also be helpful to understand the effect of any extreme events on air pollution of any study area.

The unlocking process has already begun and life has started to limp towards 'normalcy', but human reckless tendency may again give rise to the air pollution level. Therefore, a serious change in policy is needed to be enacted to keep these changes in air pollution levels into a permanent one. The government should encourage commuters to use gas-powered cars, electrifying transport, and motivate them to use more and more public transportation with proper social distancing norms rather than using private vehicles.

## 5. Conclusion

This COVID-19 lockdown period has helped us to understand the impact of human activities such as traffic movement and emission-based work in our environmental air pollution level. The present study not only emphasis on prediction of PM2.5 but also shows the spatial representation of PM2.5, which shows the changes in level due to various phases of COVID-19 lockdown. The results show that the hybrid method i.e. (MLR+ANN) performs the best with highest accuracy for the prediction of PM2.5.

Furthermore, other machine learning can also be integrated in this method for the better prediction accuracy of PM2.5. As compared to the previous year a considerable reduction in PM2.5 is also noticed due to the lockdown imposed in the country. Serious action needs to be taken by the government to keep this change in air pollution level under the permissible limit. The methodology developed can easily be used in any other study area keeping certain limitations of remote sensing data availability into mind.

## Declarations

**Acknowledgements:** Authors are grateful to the researchers supporting project number (RSP-2020/129), King Saud University, Riyadh, Saudi Arabia, as well as would like to acknowledge Central Pollution Control Board (CPCB) for the free access to ground-level PM2.5 and MODIS GSFC NASA for free access to MODIS Terra Aerosol Optical Depth data.

### Declarations:

**Funding** - Project number (RSP-2020/129), King Saud University, Riyadh, Saudi Arabia.

**Competing Interests** - The authors declare that they have no competing interests.

**Availability of data and materials** - The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Code availability** – Both software application and custom code

**Authors Contributions - Santanu Mallik:** Methodology development, Validation, Writing-original draft, Writing-review & editing. **Shivam Soni:** Data accumulation, methodology development, Conceptualization. **Krishanu podder:** Data Analysis, python coding, Writing-review & editing. **Umesh Mishra:** Supervision, Resources, Conceptualization. **Maqusood Ahamed:** Funding acquisition, Writing-review.

## References

- Agarwal, A., Kaushik, A., Kumar, S., & Mishra, R. K. (2020). Comparative study on air quality status in Indian and Chinese cities before and during the COVID-19 lockdown period. *Air Quality, Atmosphere and Health*, (2). <https://doi.org/10.1007/s11869-020-00881-z>
- Ahmad, M., Alam, K., Tariq, S., Anwar, S., Nasir, J., & Mansha, M. (2019). Estimating fine particulate concentration using a combined approach of linear regression and artificial neural network. *Atmospheric Environment*, 219, 117050. <https://doi.org/10.1016/j.atmosenv.2019.117050>
- Amanollahi, J., & Ausati, S. (2020). Validation of linear, nonlinear, and hybrid models for predicting particulate matter concentration in Tehran, Iran. *Theoretical and Applied Climatology*, 140(1–2), 709–717. <https://doi.org/10.1007/s00704-020-03115-5>

- Chelani, A. B. (2019). Estimating PM<sub>2.5</sub> concentration from satellite derived aerosol optical depth and meteorological variables using a combination model. *Atmospheric Pollution Research*, 10(3), 847–857. <https://doi.org/10.1016/j.apr.2018.12.013>
- Chu, Y., Liu, Y., Li, X., Liu, Z., Lu, H., Lu, Y., et al. (2016). A review on predicting ground PM<sub>2.5</sub> concentration using satellite aerosol optical depth. *Atmosphere*, 7(10), 1–25. <https://doi.org/10.3390/atmos7100129>
- Dantas, G., Siciliano, B., França, B. B., da Silva, C. M., & Arbilla, G. (2020). The impact of COVID-19 partial lockdown on the air quality of the city of Rio de Janeiro, Brazil. *Science of the Total Environment*, 729. <https://doi.org/10.1016/j.scitotenv.2020.139085>
- Doreswamy, K S, H., KM, Y., & Gad, I. (2020). Forecasting Air Pollution Particulate Matter (PM<sub>2.5</sub>) Using Machine Learning Regression Models. *Procedia Computer Science*, 171(2019), 2057–2066. <https://doi.org/10.1016/j.procs.2020.04.221>
- Draxler, R. R., & Rolph, G. D. (2010). HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) model access via NOAA ARL READY website (<http://ready.arl.noaa.gov/HYSPLIT.php>). *Silver Spring, MD*, 25, 5.
- Emamgholizadeh, S., Shahsavani, S., & Eslami, M. A. (2017). Comparison of artificial neural networks, geographically weighted regression and Cokriging methods for predicting the spatial distribution of soil macronutrients (N, P, and K). *Chinese Geographical Science*, 27(5), 747–759. <https://doi.org/10.1007/s11769-017-0906-6>
- Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Estes, S. M., et al. (2013). Estimating ground-level PM<sub>2.5</sub> concentrations in the southeastern U.S. using geographically weighted regression. *Environmental Research*, 121, 1–10. <https://doi.org/10.1016/j.envres.2012.11.003>
- Joharestani, M. Z., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S. (2019). PM<sub>2.5</sub> prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere*, 10(7), 12–18. <https://doi.org/10.3390/atmos10070373>
- Knoll, L., Breuer, L., & Bach, M. (2019). Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Science of the Total Environment*, 668, 1317–1327. <https://doi.org/10.1016/j.scitotenv.2019.03.045>
- Lokhandwala, S., & Gautam, P. (2020). Indirect impact of COVID-19 on environment: A brief study in Indian context. *Environmental Research*, 188(April), 109807. <https://doi.org/10.1016/j.envres.2020.109807>
- Mahato, S., & Ghosh, K. G. (2020). Short-term exposure to ambient air quality of the most polluted Indian cities due to lockdown amid SARS-CoV-2. *Environmental Research*, 188, 109835. <https://doi.org/https://doi.org/10.1016/j.envres.2020.109835>

- Mallik, S., Bhowmik, T., Mishra, U., & Paul, N. (2020). Mapping and prediction of soil organic carbon by an advanced geostatistical technique using remote sensing and terrain data. *Geocarto International*. <https://doi.org/10.1080/10106049.2020.1815864>
- Murray, N. L., Holmes, H. A., Liu, Y., & Chang, H. H. (2019). A Bayesian ensemble approach to combine PM2.5 estimates from statistical models using satellite imagery and numerical model simulation. *Environmental Research*, 178(July), 108601. <https://doi.org/10.1016/j.envres.2019.108601>
- Rodríguez-Urrego, D., & Rodríguez-Urrego, L. (2020). Air quality during the COVID-19: PM2.5 analysis in the 50 most polluted capital cities in the world. *Environmental Pollution*, 266, 115042. <https://doi.org/https://doi.org/10.1016/j.envpol.2020.115042>
- Shahriar, S. A., Kayes, I., Hasan, K., Salam, M. A., & Chowdhury, S. (2020). Applicability of machine learning in modeling of atmospheric particle pollution in Bangladesh. *Air Quality, Atmosphere and Health*, (Sportisse 2007). <https://doi.org/10.1007/s11869-020-00878-8>
- Song, W., Jia, H., Huang, J., & Zhang, Y. (2014). A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China. *Remote Sensing of Environment*, 154, 1–7. <https://doi.org/10.1016/j.rse.2014.08.008>
- Sorek-Hamer, M., Kloog, I., Koutrakis, P., Strawa, A. W., Chatfield, R., Cohen, A., et al. (2015). Assessment of PM2.5 concentrations over bright surfaces using MODIS satellite observations. *Remote Sensing of Environment*, 163, 180–185. <https://doi.org/10.1016/j.rse.2015.03.014>
- The Hindu. (n.d.). India coronavirus lockdown. <https://www.thehindu.com/news/national/india-coronavirus-lockdown-day-1-updates-march-25-2020/article31159466.ece>. Accessed 26 June 2020
- van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research Atmospheres*, 111(21), 1–10. <https://doi.org/10.1029/2005JD006996>
- Ventura, L. M. B., de Oliveira Pinto, F., Soares, L. M., Luna, A. S., & Gioda, A. (2019). Forecast of daily PM 2.5 concentrations applying artificial neural networks and Holt–Winters models. *Air Quality, Atmosphere and Health*, 12(3), 317–325. <https://doi.org/10.1007/s11869-018-00660-x>
- Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. *Geophysical Research Letters*, 30(21), 2–5. <https://doi.org/10.1029/2003GL018174>
- WHO – World Health Organization (2020). (2020). WHO – World Health Organization (2020). WHO Director-General's opening remarks at the media briefing on COVID19 - 24 March 2020. [https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-25mar2020.pdf?sfvrsn=abe86e92\\_2](https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-25mar2020.pdf?sfvrsn=abe86e92_2). Accessed 3 June 2020

Yazdi, M. D., Kuang, Z., Dimakopoulou, K., Barratt, B., Suel, E., Amini, H., et al. (2020). Predicting fine particulate matter (PM<sub>2.5</sub>) in the greater london area: An ensemble approach using machine learning methods. *Remote Sensing*, 12(6). <https://doi.org/10.3390/rs12060914>

You, W., Zang, Z., Zhang, L., Li, Y., & Wang, W. (2016). Estimating national-scale ground-level PM<sub>2.5</sub> concentration in China using geographically weighted regression based on MODIS and MISR AOD. *Environmental Science and Pollution Research*, 23(9), 8327–8338. <https://doi.org/10.1007/s11356-015-6027-9>

Zha, Y., Gao, J., Jiang, J., Lu, H., & Huang, J. (2010). Monitoring of urban air pollution from MODIS aerosol data: Effect of meteorological parameters. *Tellus, Series B: Chemical and Physical Meteorology*, 62(2), 109–116. <https://doi.org/10.1111/j.1600-0889.2010.00451.x>

Zhao, R., Gu, X., Xue, B., Zhang, J., & Ren, W. (2018). Short period PM<sub>2.5</sub> prediction based on multivariate linear regression model. *PLoS ONE*, 13(7), 1–15. <https://doi.org/10.1371/journal.pone.0201011>

Zheng, Y., Zhang, Q., Liu, Y., Geng, G., & He, K. (2016). Estimating ground-level PM<sub>2.5</sub> concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. *Atmospheric Environment*, 124, 232–242. <https://doi.org/10.1016/j.atmosenv.2015.06.046>

## Figures

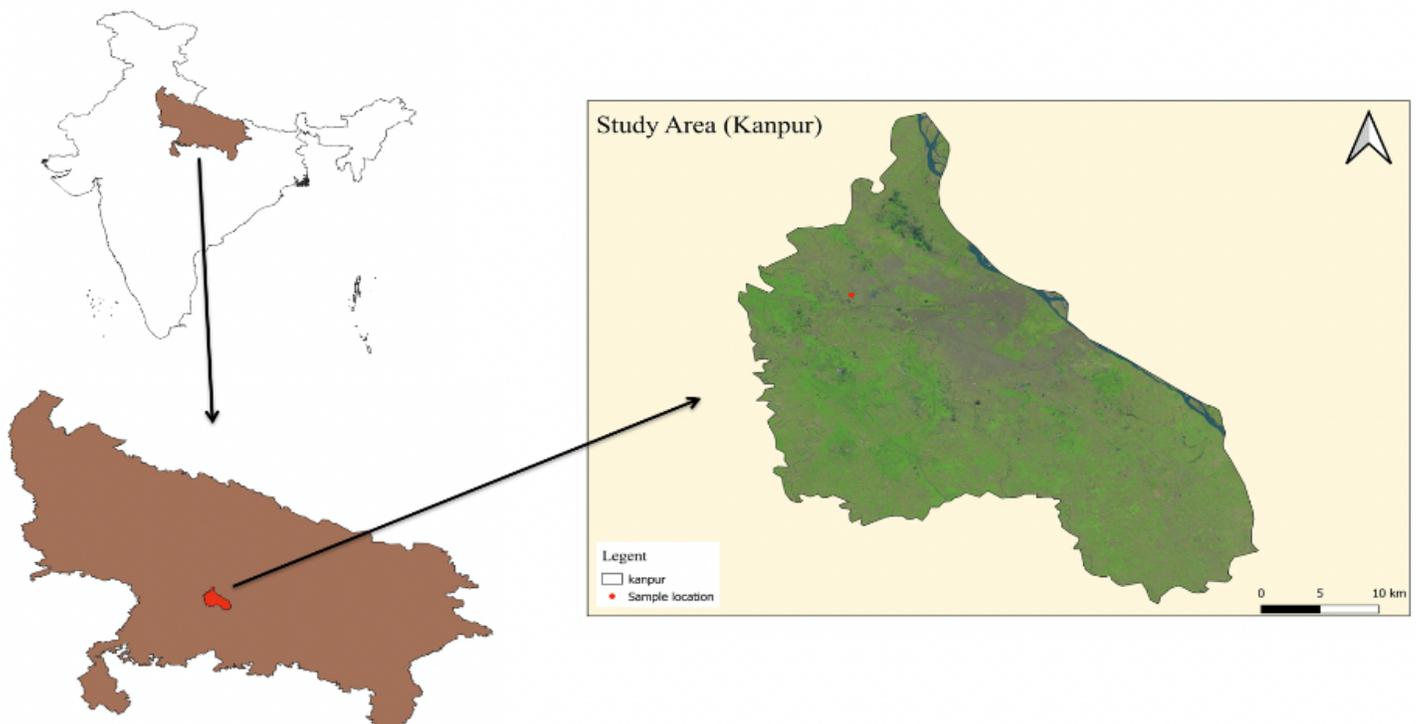


Figure 1

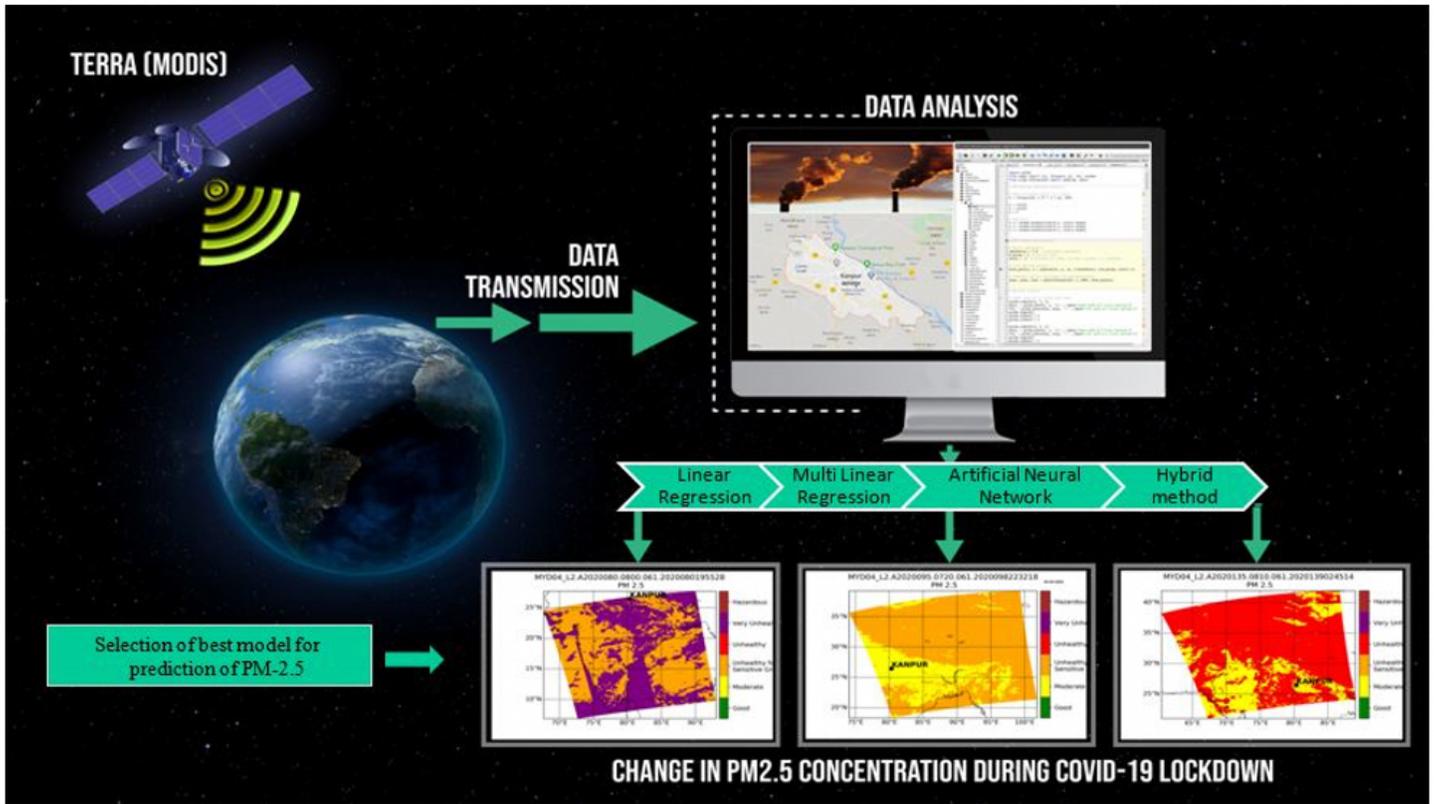


Figure 2

Flowchart & graphical representation of work

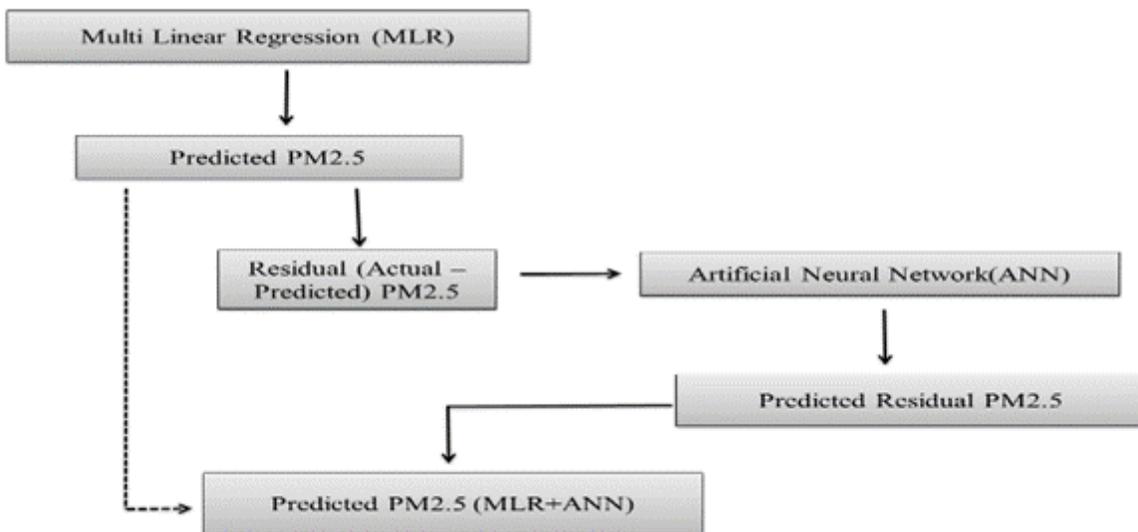


Figure 3

Schematic diagram of Hybrid (MLR+ANN) method

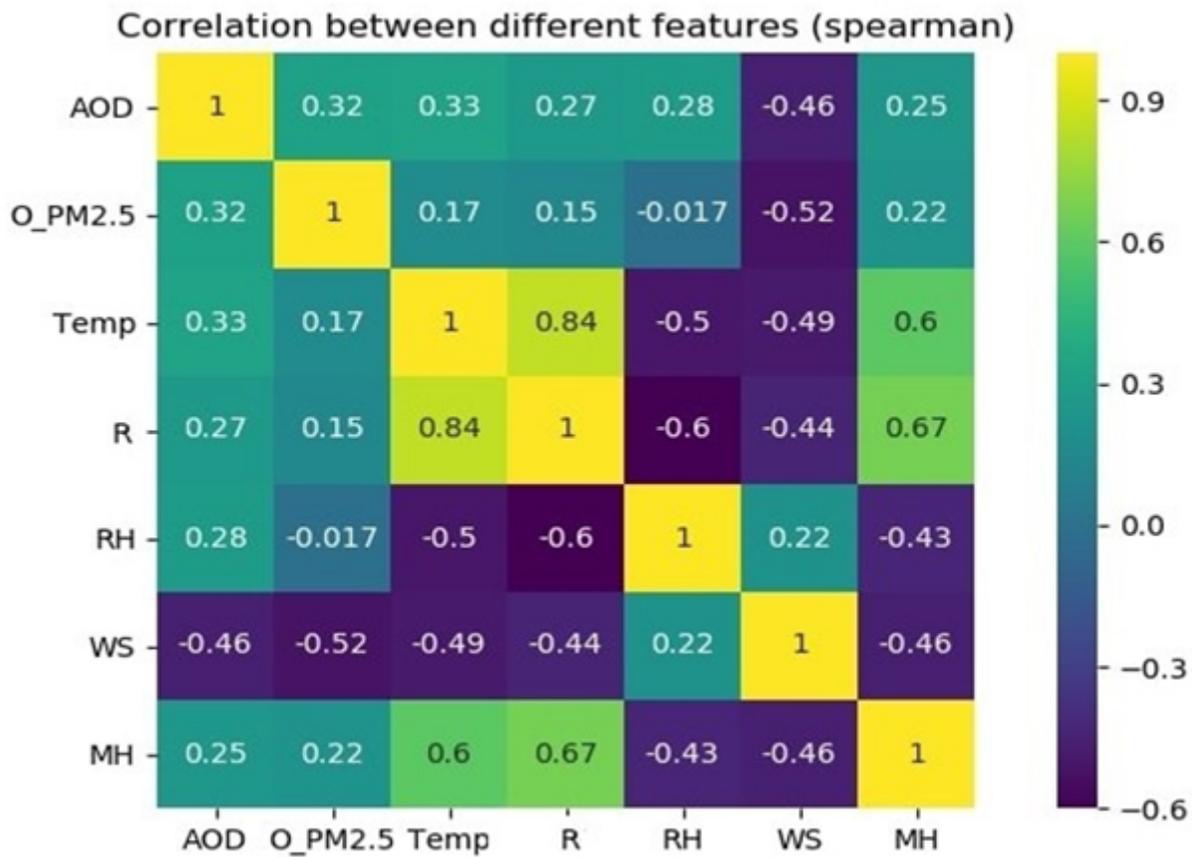
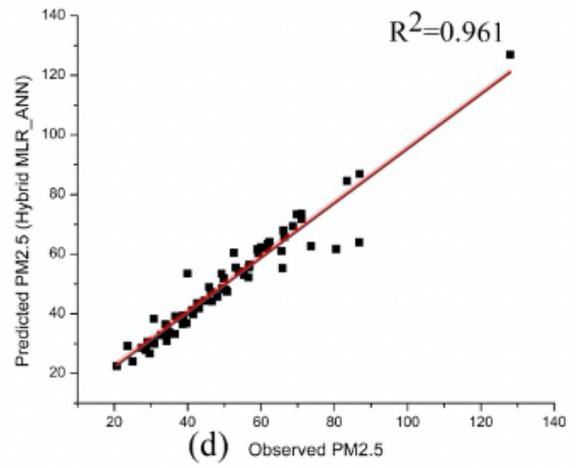
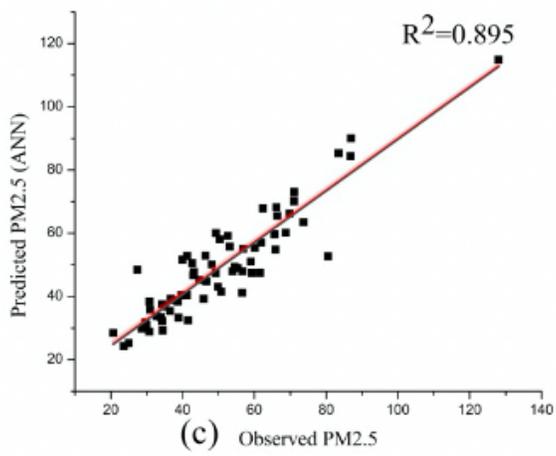
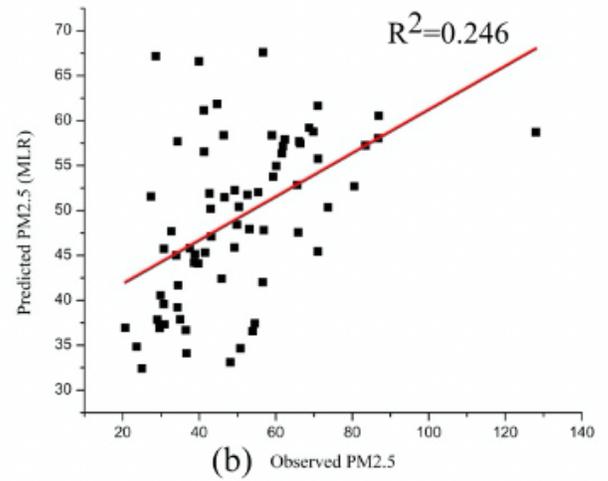
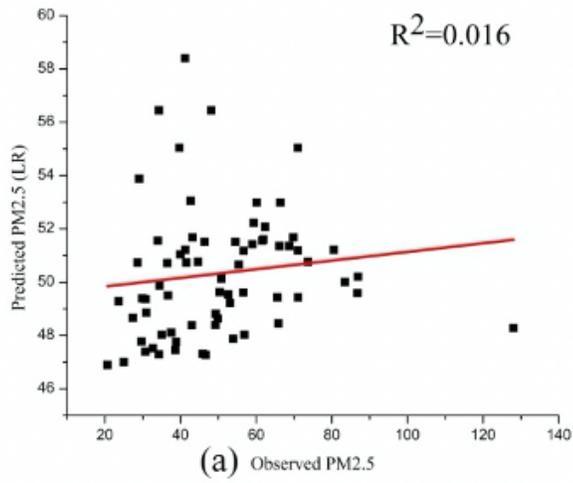


Figure 4

Spearman correlation between all the secondary variables



**Figure 5**

Prediction model comparison (a) Linear regression (LR) (b) Multi linear regression (MLR) (c) Artificial neural network (ANN) (d) Hybrid (MLR+ANN)

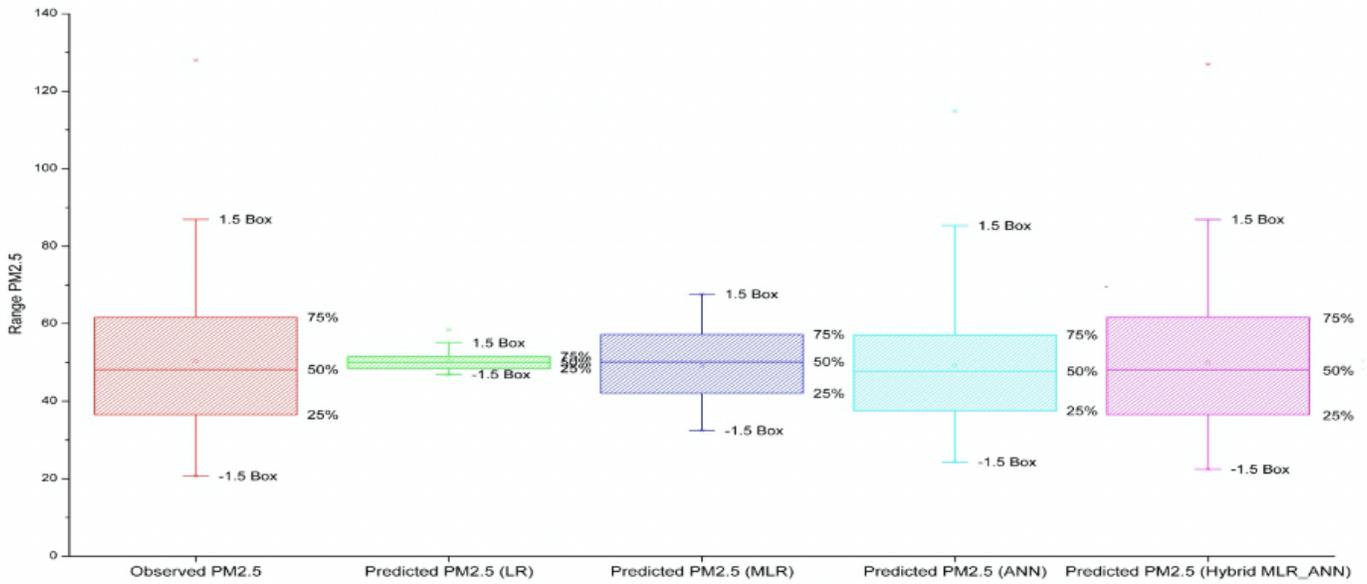


Figure 6

Box plot of the observed and predicted PM2.5 values using different methods

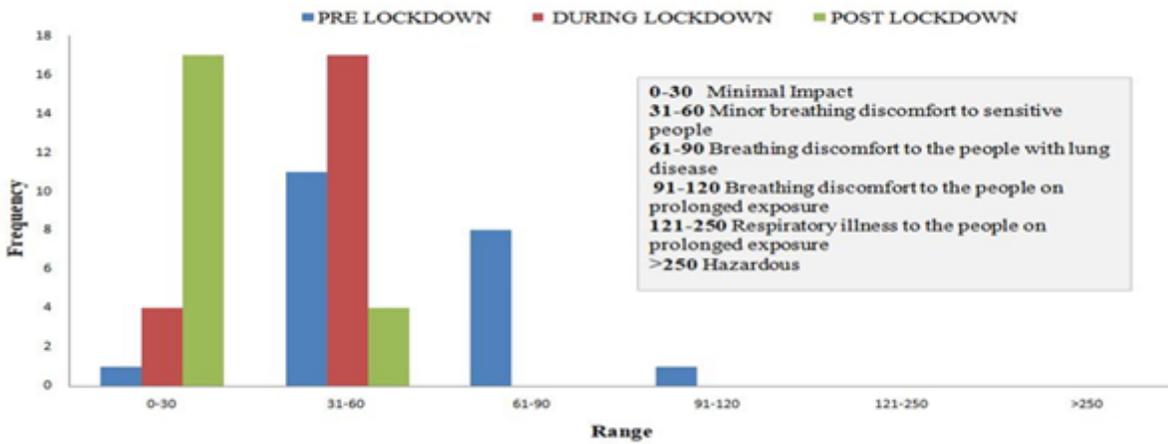
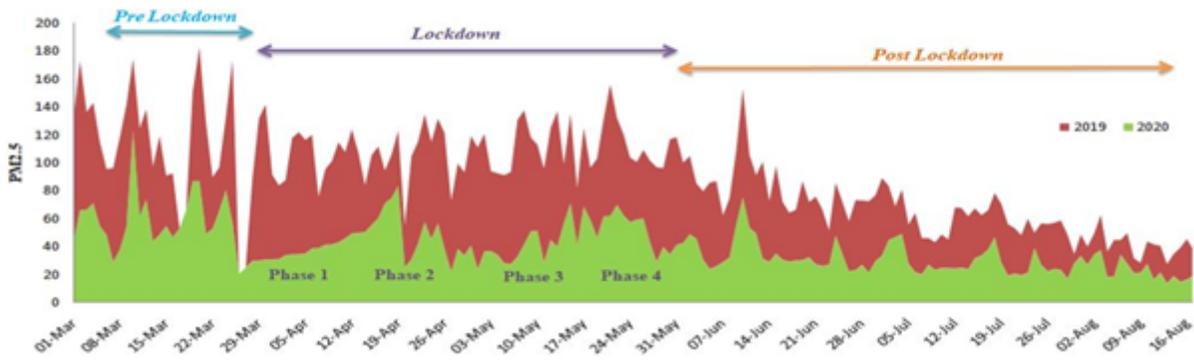


Figure 7

Histogram showing comparison of PM2.5 concentration



**Figure 8**

PM2.5 comparison with last year (2019) during various stages of Lockdown

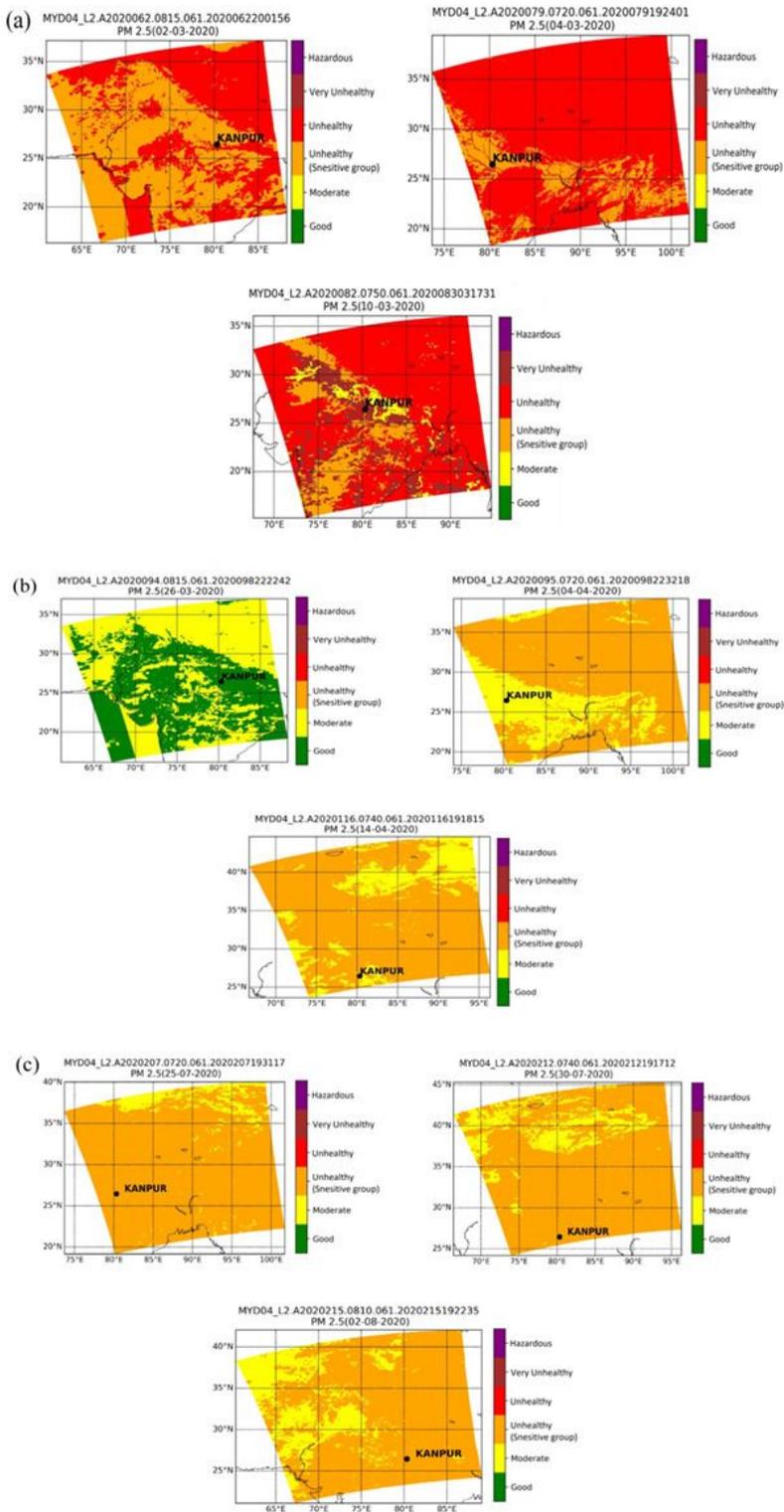


Figure 9

PM<sub>2.5</sub>concentration (a) Pre-Lockdown (b) Lockdown (c) Post lockdown