Report – Project Celestini

Raghav Bansal, Nayanika Biswas, Raman Kumar and Sahil
Delhi Technological University, virgoraghavbansal@gmail.com, nayanikab20@gmail.com, raman.dtu.a7@gmail.com, wrk.sahil97@gmail.com

Abstract - **Pollution is one of the biggest global killers, affecting over 100 million people. Fuel combustion from motor vehicles is one of the major source of pollutants, particularly particulate matter and oxides of nitrogen. 40-65% of NOx emissions is contributed by road transportation across all continents and with increasing population and urbanisation this share is increasing exponentially. There is an urgent need to study the relation of pollutant concentrations with variations in traffic to take precautionary and corrective measures to reduce contribution of vehicular traffic to ambient air pollution. Currently, insufficient sporadically located live air quality measuring stations collect pollutant concentrations. Our focus is to find relation between NOx concentrations and vehicular traffic in Delhi and design an android application that predicts NOx concentrations along roads in real time using API calls to acquire average traffic time of chosen road coordinate.**

Index Terms: ambient air pollution, android application, machine learning, oxides of nitrogen, pattern analysis, pollutant concentration of previous day, traffic time.

INTRODUCTION

Pollution is one of the biggest global killers, affecting over 100 million people. That's comparable to global diseases like malaria and HIV. 4.2 million Deaths occur every year as a result of exposure to ambient (outdoor) air pollution. [1] Fuel combustion from motor vehicles (e.g. cars and heavy duty vehicles) is one of the major source of outdoor air pollution. The pollutants with the strongest evidence of health effects are particulate matter (PM), ozone (O3), nitrogen dioxide (NO2) and sulphur dioxide (SO2). Nitrogen dioxide and nitric oxide are referred to together as oxides of nitrogen (NOx). NOx gases react to form smog and acid rain as well as play a central role in the formation of fine particles (PM) and ground level ozone. Long term exposure can decrease lung function, increase risk of respiratory conditions and increases the response to allergens. High levels of NOx can have a negative effect on vegetation, including leaf damage and reduced growth, making NOX the most harmful pollutant in the air. 40-65% of NOx emissions is contributed by road transportation across all continents and with increasing population and urbanisation this share is exponentially increasing.

The group of people most impacted by ambient air pollution due to NOx are daily commuters, particularly non – air conditioned public transport passengers, two wheelers and pedestrians. The primary step towards controlling NOX emissions is continuous monitoring of its emissions from various sources, both temporally and spatially and bringing to notice its alarming consequences to groups of people most affected by it, in a form easily understandable and decipherable by them to ensure maximum impact.

Real time monitoring stations have been set up by central authorities in India for continuous monitoring of four ambient air pollutants viz., Sulphur Dioxide (SO2), Oxides of Nitrogen as NO2, Suspended Particulate Matter (SPM) and Respirable Suspended Particulate Matter (RSPM/ PM10), however they are sparse in number as well as produce erroneous and inaccurate data. Their maintenance is often expensive and elaborate. There are only 28 live monitoring stations in Delhi including all government bodies, CPCB, IMD and DPCC. The average distance between two stations is twenty kilometers. The data acquired from these stations is extrapolated to predict air quality across Delhi. This method however popularly used, fails to predict local variations in pollutant concentrations accurately. Moreover, concentration of pollutants emitted by vehicles are significantly higher on major roads as compared to locations where most monitoring stations are situated.

Alternate methods are required which tackle all these problems and meet all aforementioned objectives simultaneously to provide consistent data which is both spatially and temporally dense.

The average time taken by a vehicle to travel between two coordinates on a road due to traffic conditions can be obtained through API. This is called the traffic time. The goal of this paper is to find relation between NOx concentrations in ambient air and traffic time so that values of NOx can be predicted using traffic parameters as the major variable. This would lead to prediction of NOx concentrations at all road coordinates where traffic time is available, thereby localizing data acquisition along roads. This will lead to greater accuracy in data and reduce absolute dependency on air concentration monitoring stations.

Delhi is chosen as the primary location for data collection and testing, specifically locations along the Ring Road with high vehicular traffic and ambient air quality monitoring stations – Punjabi Bagh, AIIMS, Mall Road(DU), Azadpur.

PREVIOUS WORK

Xia and Shao [2] in their work propose a Lagrangian model for the simulation of traffic flow on a complex road network for Hong Kong Island and use the simulated traffic flow as the basis for the estimation of traffic on the island. Furthermore using empirical emission factors for a number of vehicle categories, the emission rates of major air pollutants, CO, NOx and PM10, are estimated and the predicted emission rates are compared with measurements for several air quality monitoring stations. Hu et al. [3] find linkages between urban air quality, meteorological and traffic parameters using mutual information index and conclude that combined effect of traffic restrictions and suitable meteorological conditions are more effective in the removal of fine particles, CO and NO2.

PROPOSED ALGORITHM

I. *Initial Approach*

During the preliminary stage of the internship we tried developing a road traffic simulation model that would provide the number of vehicles present on a given section of road when provided with input parameters like free flow speed, traffic speed, traffic density and number of lanes on the road. Tail pipe emissions of these vehicles would be calculated according to standards proposed by MOBILE 6 model and air dispersion model would be applied to these emissions to map their dispersion from that particular stretch of road into ambient air.

Traffic speed is obtained using TOMTOM API which provides 4 traffic parameters of the chosen section of the road (400 meters), namely free flow speed, free flow time, current traffic speed, current traffic time. The free flow time and traffic time refer to the time taken by a vehicle on an average to cross the section of the road at no congestion and current traffic conditions respectively.

However this idea was soon discarded due to several drawbacks as mentioned below:

- Lack of uniformity in road width and posed a serious problem in lane estimation as the model required lane count of every road included in the model.
- Lack of uniformity in driving pattern in Delhi created problem in calculating traffic density across a fixed section of the chosen road. We tried solving this problem by considering 3 classification of traffic, namely free flowing traffic, medium traffic and congestion, according to traffic speed and assigning a corresponding traffic density per lane per category, on the given section. However due to combined effect of erratic lane count and non-uniform driving pattern the calculated traffic density differed from actual vehicle density on the road as calculated by camera feed on several occasions.
- For implementation of MOBILE 6 Model, after obtaining the vehicle count of a section, classification of vehicles on the basis of vehicle types, engine type, fuel type and year and manufacture is required. Such classifications are impossible to acquire in real time and assumptions would grossly prove to be erroneous.
- Exact contribution of vehicular emission to ambient air pollution concentration in not known and varies from locality to locality, so there was no method of validation of calculated values.

II. Alternate Approach

Instead of all of Delhi, the Ring Road is chosen as the focus of study. Major sections of the road is chosen which satisfied two conditions:

- Situated near a live pollution monitoring station
- Away from industrial areas

These conditions are chosen so as to eliminate two of the above drawbacks. The ring road is of uniform width throughout and hence the traffic speed can be closely related to traffic density. It is also found out that the two major contributors to air pollution are industrial areas and vehicles, so choosing a section of road near a monitoring station and also away from industrial area ensures that the data acquired from these stations provides localized signals of the effect of traffic conditions of the pollutant concentrations.

11 such locations are chosen:

- Delhi Gate
- ITO
- Ashram
- ISBT
- AIIMS
- Dhaula Kuan
- Naraina
- Raja Garden
- Punjabi Bagh
- Azadpur
- Mall Road

Punjabi Bagh is chosen as the major focus of analysis followed by three more sites: AIIMS, Azadpur and Mall Road.

Pollutant concentration dataset of Punjabi Bagh for the month of April, 2019 is chosen for chiefly 4 pollutants – Carbon monoxide, NOx, P.M. 10 and P.M 2.5. This period is chosen on the recommendation of Dr. S. Dey. Atmospheric Science Department, IITD. According to Dr. S. Dey, both meteorological and traffic parameters affect the pollutant concentrations during summer season, typically observed between mid-March to June in Delhi, but effect due to meteorological parameters dominate traffic parameters as temperature starts decreasing. The time interval between two consecutive data points is 15 minutes.

However after further analysis, as shown in the Analysis section, it is observed that vehicle emissions mainly contribute to NOx concentrations and provide the most accurate indication of vehicle emissions. On the other hand PM10 and PM2.5 are effected by several factors besides vehicle emissions. One significant instance being, movements of vehicle often significantly re-suspend the settled particulate matter rather than emitting them. The results of the analysis are discussed in the Analysis section.

Ocak and Turalioglu[4] show that there are relationships between air pollutants and meteorological factors and previous day’s air pollutants concentration. Once the relationship between various pollutant concentrations, meteorological data and traffic speed and time for Punjabi Bagh have been analyzed, we implement machine learning models to train data to predict next interval's values of NOx with these variables as features. Six models, categorized on the basis of features selected and training algorithm, are implemented. Model on the basis of features are:

- Previous interval's NOx concentration and current interval CO, NO, NO2 concentrations.
- Previous interval's NOx and current interval's CO, NO, NO2, Meteorological parameter (wind speed, wind direction, relative humidity).
- Previous interval NOx concentration and current interval Traffic Time.

Two machine learning algorithms, Linear Regression and Random Forrest, are used on each set of features, and their performance are compared under Implementation section. The solution is delivered in the form of an android applications discussed under Implementation section.

ANALYSIS

I. Diurnal Pattern

Graphs are plotted to observe the diurnal pattern. It is observed that mainly two daily peaks are seen at the same time throughout the dataset, usually between 7-9 AM and after 6:30 PM to 11 PM IST. These peaks can be closely related to office rush hours and hence can be attributed to vehicle emissions. This helped us confirm our initial assumptions of significant relation between vehicle emissions and ambient air pollutant concentrations.

II. Seasonal Pattern

NOx concentrations have been plotted over the period of 1 year, from May, 2018 to April, 2019 and typical patterns and their causes identified.

- Generally values are observed between 0-250ppb and patterns are observed as 2 major periods of high during all days of all months. On the basis of trends the year can be divided into 4 categories:
    - March-June
    - July - mid September
    - mid-September- November
    - December- February
- March- June can also be identified as the summers in Delhi and the peaks are dominant in the heat map, reaching values between 280- 300pbb.
- July –mid September can be identified as monsoons in Delhi and the rain causes settling of particles on the ground leading to lower concentrations of NOx at heights, i.e. 10 m and above, where the monitoring station sensors are present. The peaks are not very distinct.
- Mid–September–November show high average concentrations of NOx which can be attributed to crop burning in the west, in Punjab parts of Haryana which due to winds get dispersed into Delhi.

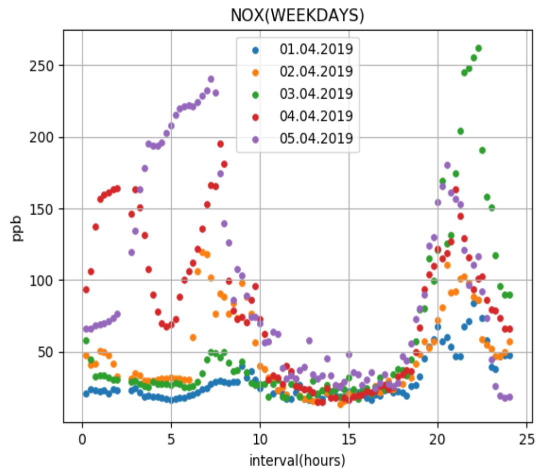

FIGURE 1
SCATTER PLOT OF NOX CONCENTRATIONS ON WEEKDAYS OF A WEEK

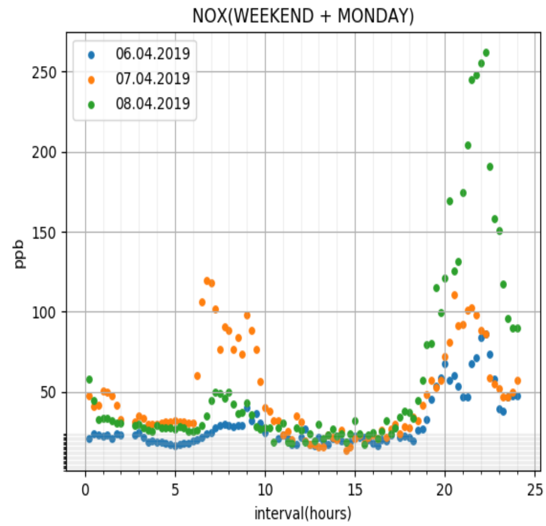

FIGURE 2
SCATTER PLOT OF NOX CONCENTRATIONS ON WEEKDAY AND MONDAY

- December- February can be identified as winters and also show high average concentrations but this is due to low temperatures which prevent pollutants from getting dispersed. The peaks are observed at the same time however have an average value between 300-400ppb.

III. Traffic Time vs NOx Concentrations

Traffic data was collected at all 11 mentioned locations over the period of seven days from July 6, 2019 to July 13, 2019 using an automated script. The Traffic time and NOx concentrations at Punjabi Bagh is plotted and a proportional (linear) relationship is obtained further strengthening our assumption that vehicle emissions significantly contribute to NOx concentrations.

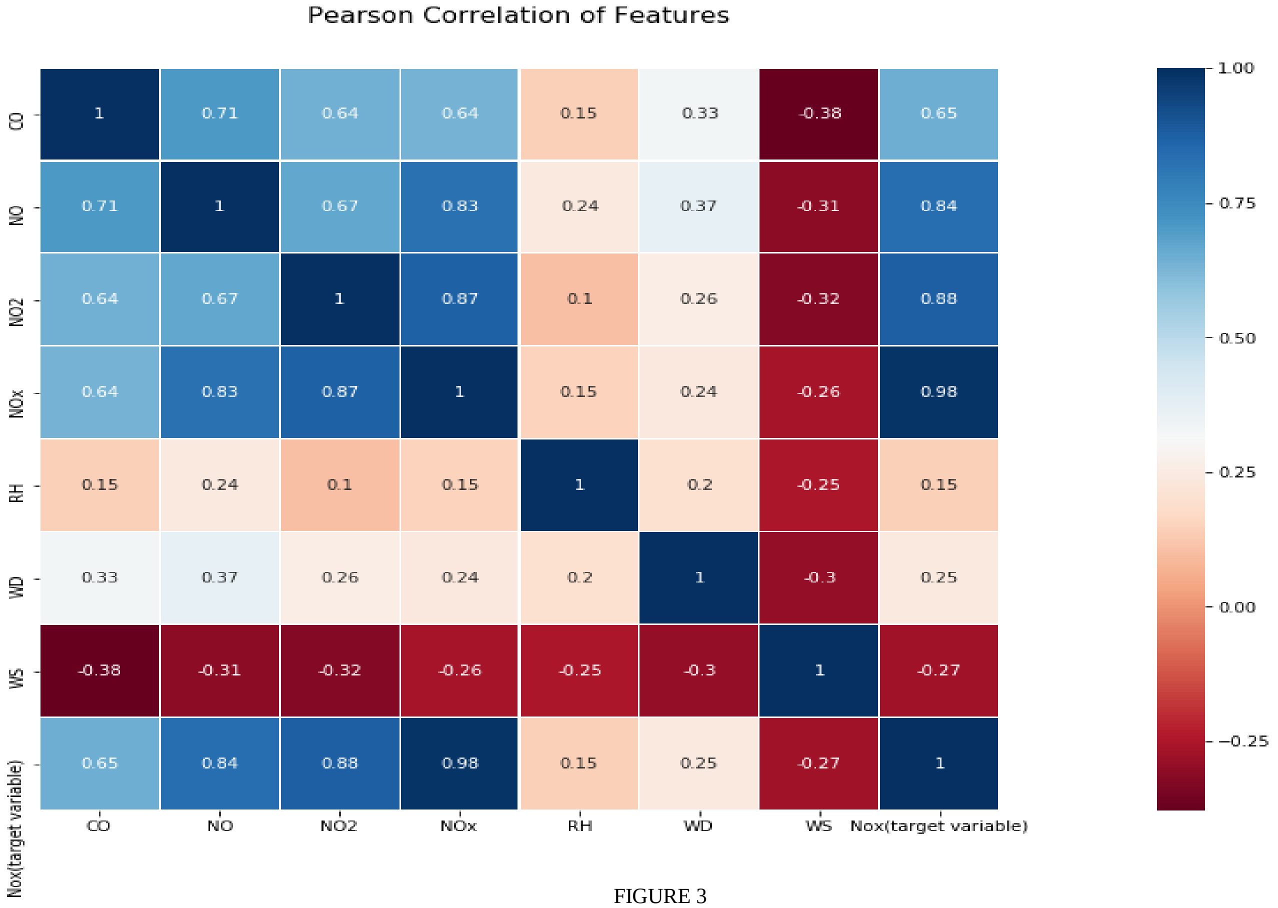

FIGURE 3
CORRELATION BETWEEN NOX CONCENTRATIONS, OTHER POLLUTANTS AND METEOROLOGICAL PARAMETERS

TABLE 1
R^2 VALUES FOR MACHINE LEARNING MODELS FOR ALL LOCATIONS

Location	Technique Used	Without Meteorological Data (R^2)	With Meteorological Data (R^2)	With Traffic Data (R^2)
Punjabi Bagh	Linear regression	0.9737267468780912	0.9719579663523836	0.9483477977693113
	Random forest	0.9831442476122269	0.9806334073905709	0.9363194232141465
Azadpur	Linear regression	0.959192360069323	0.9640617621601013	0.91504036682073
	Random forest	0.9775279626544688	0.986291097490938	0.890099411726174
R.K Puram	Linear regression	0.9660906200532701	0.9573375000044088	0.9748135452676108
	Random forest	0.9808327628469522	0.9728884437036458	0.9644246626619181
DU (Mall Road)	Linear regression	0.9840286853627477	-	0.942653817030201
	Random forest	0.9898379349086555	-	0.9280388968430056

IV. NOx concentration vs Other pollutant and meteorological parameters

Correlation between NOx concentration and other pollutant concentrations including meteorological factors were calculated. It was observed that there was a positive correlation with CO (carbon monoxide), NO2 (nitrogen dioxide) and NO(nitrogen oxide) as expected due to the fact that all these pollutants are emitted during fuel combustion. Negative correlation was observed with Wind Speed (WS)

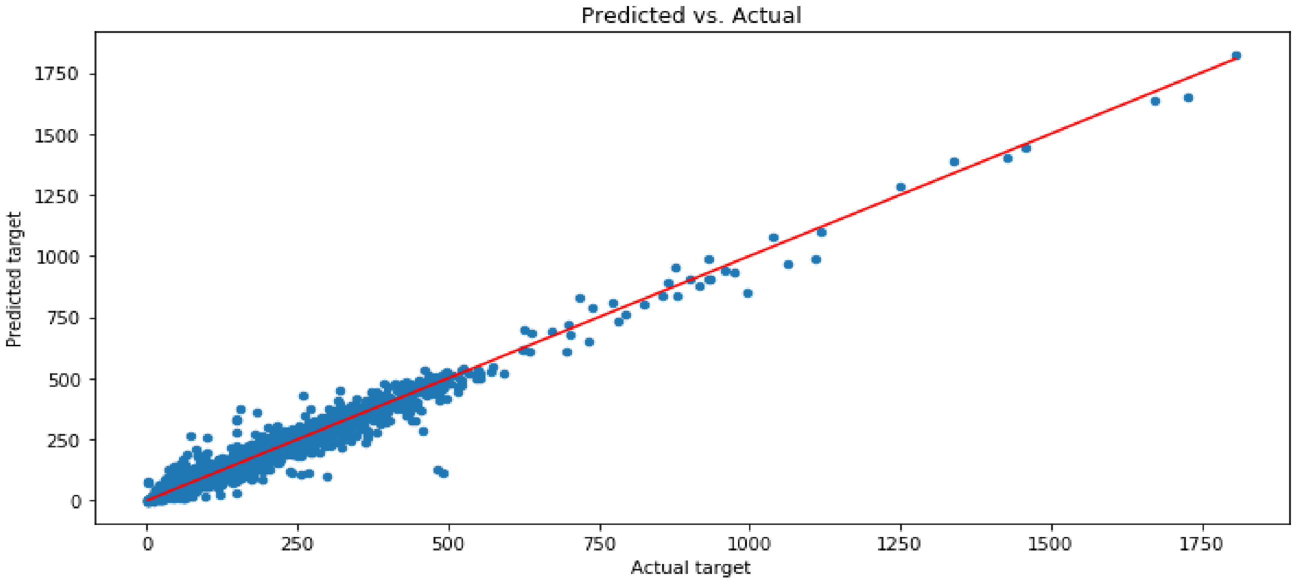

FIGURE 4
RESIDUAL PLOT: RANDOM FORREST FOR PUNJABI BAGH

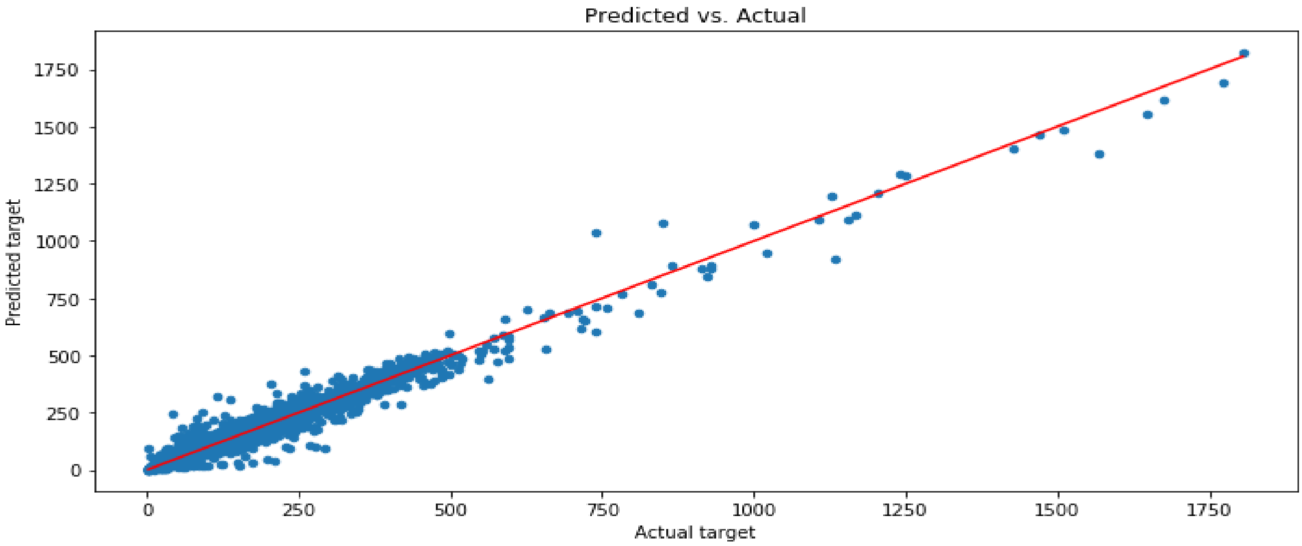

FIGURE 5
RESIDUAL PLOT: LINEAR REGRESSION FOR PUNJABI BAGH

proving that during wind disperses the pollutant present in air, thereby reducing the concentrations in air. Weak correlation is observed with relative humidity.

IMPLEMENTATION

It is concluded that NOx concentrations are dependent on traffic time and speed after similar analysis is performed for three more locations: Mall Road(DU), AIIMS and Azadpur. After establishing relationships, the six machine learning models are trained using 70:30 ratio for training and test set.respectively. The R^2 values for all the models for all 4 locations are as shown in table 1. 90% accuracy is obtained in all the three machine learning approaches. One interesting point to note here is that in the first and second approach four and six features are used to predict the label i.e. the future value of NOx. Whereas in the third approach we are only using two features (the current traffic congestion time and current value of NOx) for prediction. This shows that the correlation which we obtained between traffic time and corresponding value of NOx is accurate, reliable and sufficient enough to predict the next value of NOx. R squared value is used to determine the accuracy of our model. The residual plots are plotted to further show that the R squared values are accurate. We choose to implement the model which used random forest algorithm and Traffic time and previous NOx concentrations as features as it provides better results as compared to the model using linear regression but same features.

An android application is designed to provide users with current pollution status at the location coordinates of the device. The application functions by initially requesting for the current location of the user to select the nearest pollution station from the database of live monitoring stations. It simultaneously retrieves traffic data using the TOMTOM API for the current location and stores it in the same database. The database is stored in cloud using Firebase. Machine learning algorithm runs locally on the device to predict the NOx values of the next interval (of 15 minutes). Health effects and prevention solutions are provided to the user to reduce impact due to pollutant concentrations. An android application is chosen as a solution due to its ease of usability, customisation, accessibility and providing updates.

FUTURE WORK

The current analysis is based on datasets acquired from CPCB and is performed for only four major locations along the Ring Road in Delhi thereby making the proposed algorithm region specific. CPCB data contains several missing data point and since the model relies on the dataset for accuracy, our goal is to further validate the dataset and detect outliers by installing pollution sensors nearer to the road to make the data more reliable and consistent. Traffic data can be collected only in real time and hence the dataset for the model using traffic parameters is restricted to a period of one week. Traffic data will be collected for longer durations at all seasons to verify the generalisation of relationship between traffic parameters and NOx concentrations observed. Current scope of research is restricted to Delhi so we aim to study the traffic – pollutant concentration relationship for different parts of India where live monitoring stations are present to make our solution

ACKNOWLEDGMENT

This research was supported by Marconi Young Scholar Society and Indian Institute of Technology Delhi. We would like to thank Dr. Aakanksha Chowdhery for her continuous guidance and knowledgable insights as our mentor throughout the process. We would like to show our gratitude to Dr. Sagnik Dey and Dr. Anuradha Shukla, esteemed industrial experts in the domain of Atmospheric Sciences and Road traffic simulations, for their vested interest and enthusiastic support and guidance. This research would not have been possible without the exceptional support from Dr. Brajesh Lal, Dr. Prerana Mukherjee, Mr. Vinay Kaushik Mr. Dilip Jana, our supervisors and mentors at IITD. We are grateful for this opportunity to all.

REFERENCE

[1] World Health Organization. "Climate change and health." 2014. Web Accessed April 25, 2015

[2] L. Xia, Y. Shao, Modelling of traffic flow and air pollution emission with application to Hong Kong Island, Environmental Modelling & Software 20 (2005) 1175–1188.

[3] Hu Dongmei, Wu Jianping, Tian Kun, Liao Lyuchao, Xu Ming, Du Yiman, Urban air quality, meteorology and traffic linkages: Evidence from a sixteen-day particulate matter pollution event in December 2015, Beijing, Journal Of Environmental Sciences 59 (2017) 30 – 3 8.

[4] Ocak Sevda,Turalioglu F. Sezer, Effect of Meteorology on the Atmospheric Concentrations of TrafficRelated Pollutants in Erzurum, Turkey, J. Int. Environmental Application & Science, Vol. 3 (5): 325-335 (2008)

[5] Jereb Borut, Kumperščak Samo et al. The Impact of Traffic Flow on Fuel Consumption Increase in the Urban Environment, FME Transactions (2018) 46, 278-284.

[6] Pandey Apoorva, Venkataraman Chandra, Estimating emissions from the Indian transport sector with on-road fleet composition and traffic volume, Atmospheric Environment 98 (2014) 123-133

[7] Elminir HK, Dependence of urban air pollutants on meteorology. The Science of the Total Environment (2005), 350, 225-237

[8] Aneja VP, Agarwal A, Roelle PA, Philips SB, Tong Q, Watkins N, Yablonsky, Measurements and analysis of criteria pollutants in New Delhi, India. Environment International (2001), 27, 35-42

[9] Hargreaves PR, Leidi A, Grubb HJ, Howe MT, Mugglestone MA, Local and seasonal variations in atmospheric nitrogen dioxide levels at Rothamsted, UK, and relationships with meteorological conditions. Atmospheric Environment (2000), 34, 843-853

[10] Mohan Dinesh,Tiwari Geetam, Goel Rahul et al. Evaluation of Odd–Even Day Traffic Restriction Experiments in Delhi, India, Transportation Research Record 2627

[11] Goel Rahul, Guttikunda Sarath K., Evolution of on-road vehicle exhaust emissions in Delhi, Atmospheric Environment 105 (2015) 78-90.

[12] Cogliani Euro, Air pollution forecast in cities by an air pollution index highly correlated with meteorological variables, Atmospheric Environment 35 (2001) 2871-2877.

[13] Brian Y. Kim, Roger L. Wayson, and Gregg G. Fleming, Development of Traffic Air Quality Simulation Model, Transportation Research Record 1987

[14] User's Guide to MOBILE6.1 and MOBILE6.2: Mobile Source Emission Factor Model. EPA Report EPA420-R-02-028. Assessment and Standards Division, Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Washington, D.C., Oct. 2002.

[15] 1. Kim, B. Y. Predicting Air Quality Near Roadway Intersections Through the Application of a Gaussian Puff Model to Moving Sources. PhD dissertation. Department of Civil and Environmental Engineering, University of Central Florida, Dec. 2004.

[16] GU et al.: Recurrent Air Quality Predictor Based On Meteorology-And Pollution-Related Factors, IEEE Transactions On Industrial Informatics, Vol. 14, No. 9, September 2018

[17] Nagendra S. M. Shiva, Khare Mukesh,, Modelling urban air quality using artificial neural net, Clean Techn Environ Policy (2005) 7: 116–126