Complexity of Factor Analysis for Particulate Matter (PM) Data: A Measurement Based Case Study in Delhi-NCR

ISMI ABIDI, Indian Institute of Technology Delhi, India
SAGAR RAVI GADDAM*, Oplus, India
SASWAT KUMAR PUJARI*, Oracle, India
CHINMAY SHIRISH DEGWEKAR*, Eightfold, India
RIJUREKHA SEN, Indian Institute of Technology Delhi, India

Developing countries are home to the most polluted cities in the world. Particulate Matter (PM), one of the most serious air pollutants, needs to be measured at scale across urban areas in such countries. Factors potentially affecting PM like road traffic, green cover, industrial emissions etc., also need to be quantified, to enable fine-grained correlation analyses among PM and its causes. This paper presents an IoT platform with multiple sensors, latest deep neural network based edge-computing, local storage and communication support – to measure PM and its associated factors. Through real world deployments, the first in depth empirical analysis of a government enforced traffic control policy for pollution control, is presented as a use case of our IoT platform. We demonstrate the potential of IoT and edge computing in urban sustainability questions in this paper, especially in a developing region context. At the same time, we show how complex a real system like Particulate Matter's factor analyses can be, and urge environmentalists to use sensors networks and fine-grained empirical datasets as ours in future, for more nuanced and data-driven policy discussions.

ACM Reference Format:
Ismi Abidi, Sagar Ravi Gaddam, Saswat Kumar Pujari, Chinmay Shirish Degwekar, and Rijurekha Sen. 2022. Complexity of Factor Analysis for Particulate Matter (PM) Data: A Measurement Based Case Study in Delhi-NCR. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22), June 29-July 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3530190.3534808

1 INTRODUCTION

Air pollution is a serious cause of concern for modern civilization, specifically in the big cities of developing countries. It has reached life-threatening levels in Delhi-NCR, one of the largest and most densely populated urban centers in the world. In this paper, we use Delhi-NCR as a use case for scalable PM measurement and PM related policy analyses. This city is an interesting use case for PM related policy debates. Below are some examples of the policy questions discussed in the urban area, where empirical data would assist more quantitative decision making.

Green cover vs. pollution: While economic growth requires more infrastructure to be built, whether that should come at the cost of felling more trees is constantly debated. Earlier, citizens were not so conscious about the relation between green cover and air pollution, and as a result "One tree (has been) cut every hour over last 13 years, says Delhi govt data", according to news reports [46]. However, recent government notices on tree felling for housing development projects [14, 45] have been met with tremendous citizen resistance [15, 39], finally causing the government to back off [32]. Data driven analyses can assist in quantitative decisions, balancing environmental requirements with

*This work was performed when the authors were at Indian Institute of Technology Delhi sustainable infrastructure needs. Such analyses will help in mobilizing data-driven citizen resistance movements, when necessary, and aid the government in pushing through a sustainable development project, where suitable. Section 2.3 presents how to quantify PM, percentage of green cover and built-area of different regions through direct measurements, while the exploration of the relation between PM and these factors are done in Section 3.1.

Vehicle count vs. pollution: The Delhi government piloted an odd-even policy to reduce the number of vehicles and associated vehicular emissions between 01-15 Jan, 15-30 April 2016, and November 4-15, 2019. Four-wheelers with odd and even-numbered plates were allowed on alternate days. This was met with angry protests from citizens, citing that the public transport infrastructure was not ready to take on the burden of the extra commuters [27, 44]. Some citizens, on the other hand, enthusiastically co-operated [38]. The saddest reality was that, such an important policy affecting millions of citizens' daily lives, could not be conclusively evaluated in relation to PM control. Conflicting studies suggested pollution decreased [13, 43], remained the same [17, 21] and even increased [9, 23]. All studies have commented on the paucity of measurement data, not only pollution data, but also vehicular count and classification data [35]. The greatest confusion came regarding whether vehicle counts actually decreased on odd-even policy days, or in fact increased – people hired vehicles with appropriate number plates, two-wheelers were exempted from the odd-even policy rule which were used in unprecedented numbers, old buses were used to boost public transport and so on. Section 3.4 presents an empirical evaluation of the second round of odd-even policy enforcement in Nov 2019 [13, 21, 43]). The effect of vehicle count on PM, along with fine-grained classification of the different vehicle classes, with and without the policy, has been empirically studied for the first time, with our IoT devices deployed across real traffic intersections.

Local sources vs. pollution: Pollution is exacerbated from local sources like industrial emissions, agricultural crop burning, cooking fuels and diesel generators in commercial areas etc.. Policies experience the bloodiest contests when there are dire economic consequences of shutting down such local pollution sources. Fierce retaliations have been staged by thousands of workers, who lost livelihood in industrial shutdown or relocation in Delhi-NCR in 1996 and 2001 [10, 16, 26]. Similar violent resistance have been seen among farmers, who burn their crop stubble in Punjab and Haryana, neighboring states of Delhi, adding to PM concentration in Delhi-NCR. The economic cost of environmental friendly stubble disposal techniques are still beyond most farmers, according to the protesters, and the fines imposed on them for crop burning increase their existing burden of financial debts [11, 12, 25, 34, 40]. There is no clear evidence of the environmental benefits obtained through such policies, which have extreme economic repercussions on the larger sections of society. Hence, empirical measurements of PM data in the urban industrial and agricultural border areas are needed. *This paper does not directly study these local sources at city scale. But it presents an IoT platform in Section 2.1 that uses the recent promise of well-calibrated low cost PM sensors [8, 19, 28, 36, 52]. It is augmented with other sensors to collect auxiliary information for PM policy analysis. This platform and our methodology can be used, in future, for more extensive data collection of PM and its local sources. In this paper, we show a case study by deploying 26 instruments at different locations in a university campus, and 3 instruments at traffic intersections.*

Existing literature on air pollution and factor analyses [31, 41, 42, 48, 50] show different air pollutants such as PM,NOx, SOx, Ox and black carbon have different correlation with different meteorological factors for a given area. Li et al [29] shows in their research that major influencing factor for pollution changes with the study area. As a result, factor analysis for pollution conducted for one study area, such as Beijing and other cities in China in prior works, might not be valid for Delhi-NCR. Furthermore, prior studies collect data from a limited number of monitoring stations or EBAM units [41, 48, 50], which are already deployed and available in an urban area. Deploying new monitoring stations for the purpose of particular policy analyses might be prohibitively costly. Recent trends on using low cost

Work	Area	Cost	Measured Parameters	Remarks/Focus
Tiwari et al [48]	Delhi	High	Black Carbon, Particulate Matter	Seasonal variation and correlation
Srinivas et al. [41]	Delhi	High	Paticulate Matter	Air quality forecast
Yadav et al [50]	Udaipur	High	O_3,CO,NOx	Seasonal variation
Opensense [7, 30]	Zurich	Medium	O_3,UFP	Calibration error reduction
Gao et al. [19]	Xi'an	Low	Particulate Matter Meteorological parameters	Identify PM2.5 hotspot
Budde et al [3, 4]	-	Low	Particulate Matter	Calibration and sensor fusion
HazeWatch [24]	Sydney	Low	CO, NOx, Meteorological parameters	Data Modeling and visualization
AQ360 [18] ImageSensingNet [51]	Beijing	Low	Paticulate Matter, Haze pictures	AQI recognition using unmaned aerial vehicle (UAV)
This Work	Delhi	Low	Particulate Matter Meteorological Parameters, Traffic Density	Static and dynamic factor measurement and correlation, odd-even policy analysis

Table 1. Related work on air pollution sensing

sensors for environmental sensing [3, 4, 30] enables the authors in this paper to deploy additional instruments at interesting locations like traffic intersections. This helps in measuring the effect of odd-even vehicular control policy on PM. Depending solely on existing pollution measurement infrastructure, as in prior work [31, 41, 42, 48, 50], would not have made such dynamic deployment and policy analysis feasible. Furthermore, each IoT unit in this paper is augmented with other sensors to collect auxiliary information for PM policy analysis – a camera to count and classify road traffic to correlate PM with vehicular emissions, GPS to query Google satellite images for green cover information etc. Works like [7, 30, 52] focus on calibrating the low-cost sensors for real-world deployment. This paper builds on the promise of the prior work on low-cost sensing. Table 1 summarizes the relevant projects conducted on the environmental monitoring. Despite the fact that these projects replicate efforts of the air pollution monitoring, we believe they are all worthy endeavours since they investigate diverse deployment options for different pollutants in different parts of the world.

Contributions: To summarize, this paper contributes the following:

- We build a low-cost platform that uses the recent promise of well-calibrated low-cost PM sensors for air pollution sensing and auxiliary sensors like a camera for PM policy analyses.
- We investigate ways to classify the percentage of green cover and built-area in various places, as well as the relationship between PM and these variables.
- We perform correlation analysis of static and dynamic factors with particulate matter.
- We empirically study the effects of vehicle count on PM, along with the fine-grained classification of the different vehicle classes, with and without the odd-even policy.

RoadMap: In Section 2, we first describe the sensor requirements and then the design considerations of our customized platform. Next, we describe the method for classifying the green cover, built-up, residential and commercial area. In Section 3, we correlate PM with multiple static and dynamic factors. Finally, we conclude the paper in Section 4 and discuss the future directions.

2 PARTICULATE MATTER AND AUXILIARY DATASETS COLLECTION AND PRE-PROCESSING

2.1 Customized Hardware Platform Design, Prototyping and Deployment

For deployment based empirical analysis of Particulate Matter (PM) and factors affecting PM, we design and prototype our custom device. This instrument contains the following sensors.

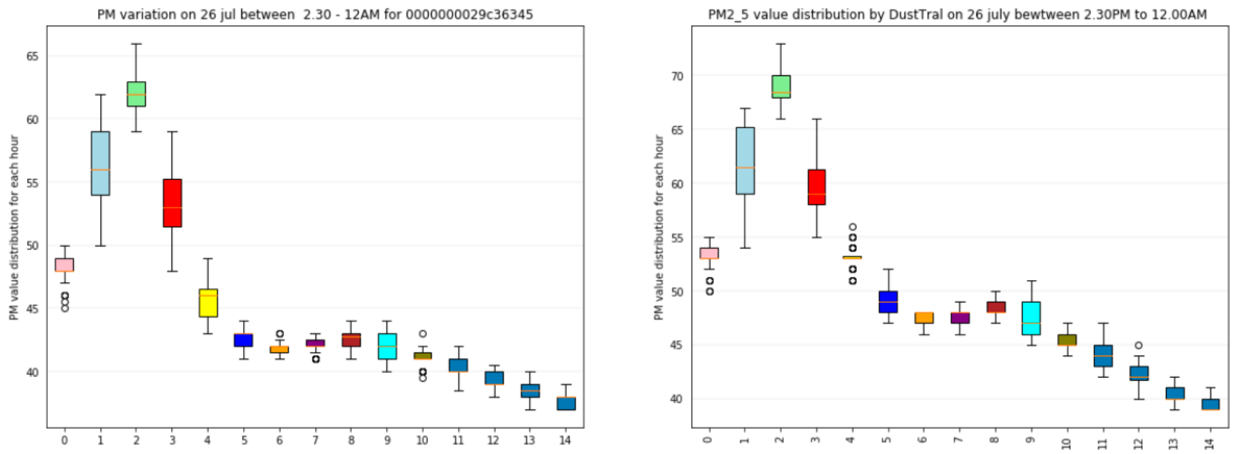


Fig. 1. Low Cost sensor's PM 2.5 values on the left vs. a co-located TSI DustTrak's PM 2.5 values on the right, on a sample day. While their costs are 30 USD and 10K USD respectively, i.e. an order of magnitude different, their measured PM values are almost identical.

(1) Low-cost, embedded PM sensor, as measuring PM is the main goal. We have computed the similarity between the low-cost PM 2.5 sensors (30 USD) and the industry grade TSI DustTrak (10K USD) for some random days. We have observed that the low-cost PM 2.5 and DustTrak have a similar distribution. Figure 1 shows the boxplot as an example of the similarity between the two instruments for a sample day.
(2) Temperature and humidity sensors are included for appropriate calibration of the PM values measured with the embedded sensors, against industry grade reference monitors. They can also be used for meterological factor correlation analysis.
(3) We have included GPS to know the location of the device in a scalable deployment. As for now, we have considered only the static deployment case and therefore, we know the GPS value of the deployed unit. However, in future, we plan to use our device for vehicular-based sensing, and hence GPS sensors are included in the prototype. GPS coordinates are also important for crawling other web data sources for offline analysis. Offline analysis can include measurement of green cover using Google satellite images, or measurement of urban composition (industry or shopping areas or residential complexes) by using Google Places.
(4) We have included a camera to take traffic images in the region of sensor deployment. The images can be used to compute vehicle counts and classifications in the device neighborhood.

Table 2 summarizes the sensors and their purposes. For communication between the deployed sensors and the backend server, cellular connectivity is the only viable option in a developing region. To reduce cellular communication costs, edge-computing on the device is exploited. The sensor data is processed locally, and the high level inferences are communicated to the back-end server.

To collect PM data, we deployed these instruments at different locations in a university campus, since Dec 2018. We also deployed the units at several traffic intersections in Delhi-NCR. We will discuss the data based empirical observations, in relation to road traffic densities, in Section 3.

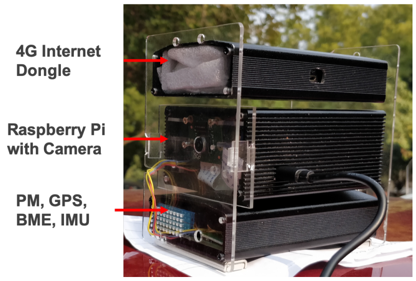


Fig. 2. Hardware platform first prototype, which was discarded for looking too attractive tempting theft at deployment sites.

Sensor	Purpose
PM	PM 2.5 and PM 10 counts
BME	temperature and relative humidity to correct PM
GPS	mounting vehicle location, air quality factor correlation
Camera	vehicle count and classification in area of static deployment

Table 2. Sensors and their uses in our custom device

Fig. 3. Hardware platform second prototype showing the sensors, micro-controller, our custom PCB etc., inside ABS plastic boxes.

Fig. 4. Hardware platform second prototype multiple units. These were deployed in university campus, at traffic intersections etc.

2.2 Design Considerations

To summarize, following are some of the challenges that need to be addressed while doing a scalable sensing.

Packaging: The packaging of the IoT unit changes based on where it is mounted, and the available fixtures. Our first prototype is shown in Figure 2. As the device looked glossy and expensive, there were concerns raised about theft by the deployment partners. Thus, we repackaged the device in a mundane looking ABS plastic box (second prototype shown in Figure 3 and Figure 4).

Calibration: Although we have seen a large similarity between our low-cost PM sensor and TSI DustTrack, we need to re-calibrate the sensor periodically for a consistent performance. Methods presented in [4, 52] can be applied to re-calibrate the PM sensors.

Local Processing: Edge computing on the device is used to cut communication expenses. Instead of sending the raw sensors data, the data are processed locally, and high-level inferences are sent to a back-end server. The IoT device stores one PM value, one GPS value, one BME value, one image. The device sends the raw PM, GPS, and BME value to the server. However, instead of sending the raw images, the device sends the vehicle count and classification results to the server. This reduces the communication cost. In such scenario, we need to balance the communication cost and computation cost.

Power consumption: Our device is not battery powered. As for long term operation, charging batteries will not be sustainable. We tested the device at 5V constant supply for several hours. Peak current was 1.2-1.5A, and hence the power remains within the 10 Watts.

Heating: The other important factor to consider for scaling up deployment is heating of the device from data processing, and possibilities of self-throttling. The average CPU usage, memory usage and SoC temperature are 50%, 11% and 55°C respectively, as measured with Linux command line utilities, for prolonged running of the software for the device. The temperature range is safe as the device self-throttles at 85°C.

In our deployments, we do not see any heating, self-throttling and related rebooting issues. The only reason where a device rebooted was a power supply interruption. However, the sense-compute-store-communicate loop automatically started once the power was restored. The intended behavior for these units is to remain operational as long as the mounting infrastructure is on and unit gets power.

2.3 Green Cover and Builtup Area Detection Using Satellite Image Processing With Neural Nets

Green cover regions and built-up regions are important spatial features that can be extracted from satellite images. We use ArcGIS's WorldImagery API to collect the satellite images of Delhi. We divide the whole area of Delhi into 500m x 500m square grid. Using the WorldImagery API, we download the image corresponding to each grid cell with appropriate location coordinates. The images have a resolution of 1m, i.e. distance between two pixels is approximately 1m on the ground. In total, we collected 9948 images for the city of Delhi.

To find out the percentage of green cover and builtup regions in any given area, we use image segmentation algorithms which classify each pixel in an image to a particular class. Deep Learning models are currently the state of the art for image segmentation tasks. Hence, we experiment with three different deep learning models: Mask-RCNN [20], Deeplab [6] and UNET [37] to calculate the percentage of green cover and builtup regions for an image accurately. We first preprocess the satellite images and annotate them for preparing the ground truth images. These images are then given as input to the deep learning models for achieving the task. We use an open source tool, namely, VGG image annotator (VIA) [1], to manually annotate the images with ground truth classes for the segmentation task. It allows creating polygons that can enclose the green cover area and buildings which can be of any shape and size. After annotations, VIA tool generated a json file having all the annotations. This json file is then processed to generate masks as per different segmentation model requirements.

For segmentation task, we first examine Mask-RCNN model [20]. Mask-RCNN performs instance segmentation in images by first performing object detection and fixing the bounding box around all objects of different classes. It then performs segmentation inside the bounding boxes to classify all the pixels of that object.

We next examine DeeplabV3+ [6], developed by Google, which has achieved 86.9 mIOU on PASCAL VOC 2012 benchmark. It has also been successfully used in segmentation of satellite images. UNET is the third image segmentation model [37] examined by us. This is a binary classification model i.e it can be used to classify two classes. As the number of classes in our problem are three (including background class), we train three different models for classifying the classes pairwise and then calculate the mean IOU by averaging the results of the three models on each image.

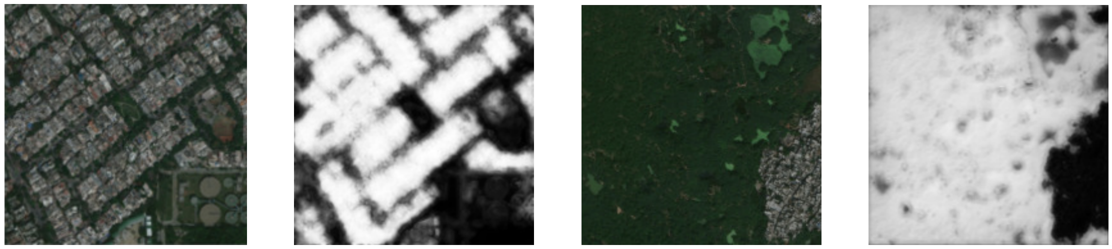
For Mask-RCNN, labeled images are divided into a train-test split of 400 and 50 images. For semantic segmentation using DeepLab and UNET, 350 images are used for training, 50 images for validation and 50 images for testing. The accuracy results are given in Table 3. We observe that Mask-RCNN has very poor accuracies. This is further visually illustrated in Fig 5. Mask-RCNN does instance segmentation in which different instances of a particular class are treated as individual objects. As a result, the model tries to learn the different shapes, sizes, symmetries and other complex features about the object and for that large amount of data must be trained for achieving good accuracies. However, in semantic segmentation only the characteristics of the pixels are learnt to classify them. Both DeepLab and UNET perform semantic segmentation, and achieve reasonable accuracies (Table 3). UNET can be trained in much less time than DeepLab, and therefore has been used in further analyses. UNET’s precise masks are visually illustrated in Figure 6.

Model	**Training**		**Validation**		**Test**	
	Number of Images	**MAP Score**	**Number of Images**	**mIOU**	**Number of Images**	**mIOU**
Mask-RCNN	400	0.24	-	-	50	0.21
DeepLab	350	0.75	50	0.68	50	0.67
UNet	350	0.73	50	0.70	50	0.69

Table 3. Accuracies obtained using different image segmentation models

(a) Original Image (b) Predicted Image

Fig. 5. Segmentation results of Mask-RCNN model, showing very imprecise masks.



(a) Original Builtup Image (b) Predicted Builtup Image (c) Original Green-cover Image (d) Predicted Green-cover Image

Fig. 6. Segmentation Results of UNET model, showing very precise masks.

2.4 Residential and Commercial Area Classification Using Google Places and OLX data

We classify a given area of Delhi as residential or commercial. *Olx.in* contains a large number of advertisements related to sale or rent of the residential and commercial spaces. These advertisements are used to classify areas as residential and commercial, by browsing residential and commercial advertisements posted in Delhi and collecting location information of each advertisement. We collect a total of 104000 ads belonging to residential locations and 24000 ads of commercial locations. We further characterize places as establishments using Google places API, a service that returns information about places around a given location. Some of the categories which are returned by the API are food, health, transport, religion, education, financial services, shopping, retail and entertainment. We collect this information for the whole of Delhi. Given a particular area of Delhi, we want to calculate the confidence level for that area to be a residential or commercial area. We select an area size of 500m x 500m i.e 250000m^2 to match with the area represented by the satellite images. For each such block, the number of residential and commercial data points are calculated based on Olx and Google places data. Google places data points with categories as food, daily needs, religion, safety, education and health services are considered as residential data points and those with categories as retail, shopping, financial services, tourism, entertainment and transport are considered as commercial data points. We have used the following formula to find the confidence level.

$$\%residential = \frac{\#residential\_data\_points}{\#residential\_data\_points \; + \; \#commercial\_data\_points} \tag{1}$$

$$\%commercial = 1 - \%residential \tag{2}$$

3 CORRELATING PARTICULATE MATTER (PM) AND FACTORS POTENTIALLY AFFECTING PM

3.1 Static Factor Analysis for PM 2.5: Green Cover vs. Built-up Areas, Commercial vs. Residential Areas

Spatial features like greencover vs. builtup area and residential vs. commercial areas, do not change rapidly in a city. We, therefore, term these factors as static factors, and examine how these factors are related to PM 2.5 values. Since we cannot deploy our custom instruments at significant granularity all over Delhi, for which we have collected the satellite images and Google and Olx datasets, for this analysis, we use PM 2.5 values from government deployed pollution monitoring stations (locations of these sensors are shown in Figure 10).

We create a feature set for each station, considering 1Kmx1Km area around each station. We calculate the percentage of green cover and built-up area using the U-Net model for each station. We further compute the percentage residential

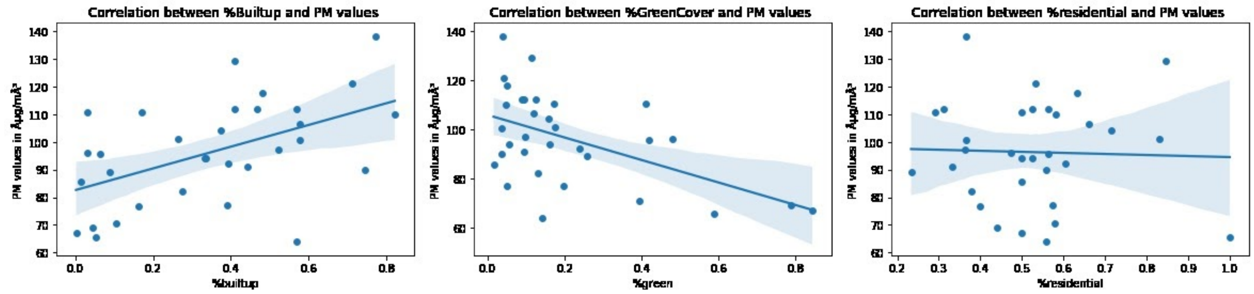


Fig. 7. Correlation plots between static factors like %builtup, %greencover and %residential with PM 2.5 for sample day 134

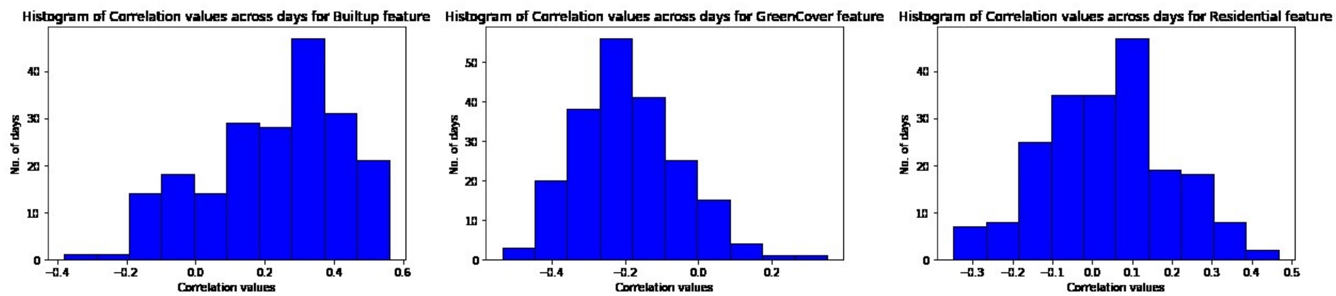


Fig. 8. Correlation histograms across 204 days, using PM 2.5 data from government deployed pollution monitoring stations

values using Equations 1 and 2. Now for a Day_i , we compute the daily median PM value for each station. We find the correlation between these PM values and %green cover for Day_i. Similarly we do it for %builtup and %residential. We plot the feature values, with the fitted lines for Day 134 (Dec 30, 2019) as shown in Figure 7. Across 204 days for which we crawl the static sensors' data, we plot the histograms of the correlation values represented in Figure 8. We observe consistent positive correlation between %builtup and PM2.5 values (avg: +0.228), and consistent negative correlation between %green and PM2.5 values (avg : -0.194). We find almost no correlation with %residential (avg : 0.034). So, %residential or %commercial do not seem to have any linear relationship with PM.

3.2 Ranking Google Places Categories In Relation To Particulate Matter

As percentage of residential or commercial feature had negligible correlation with PM values, we seek to find the importance of these Google Places based features at a more granular level.

We start with clustering the AQI monitoring stations based on PM trends throughout the year. Ideally the stations report data every 15 mins. But there are days when this frequency is less. We remove such days to select 32 stations which had 15 min frequency data for 204 days across the year. Each station is characterized by a 204 dimensional feature vector, where each value represents the median PM 2.5 value for a day. We then perform time series K-Means clustering using DTW (Dynamic time wrapping) as distance metric. Using Elbow Point Technique, the optimal number of clusters is found to be 3. Figure 9 summarizes the averages across each cluster. Cluster 1 (red line) has consistently higher PM 2.5 values compared to cluster 2 (blue line), and cluster 3 (green line) shows lowest PM. The spatial distribution of the clusters across Delhi is shown in Figure 10.

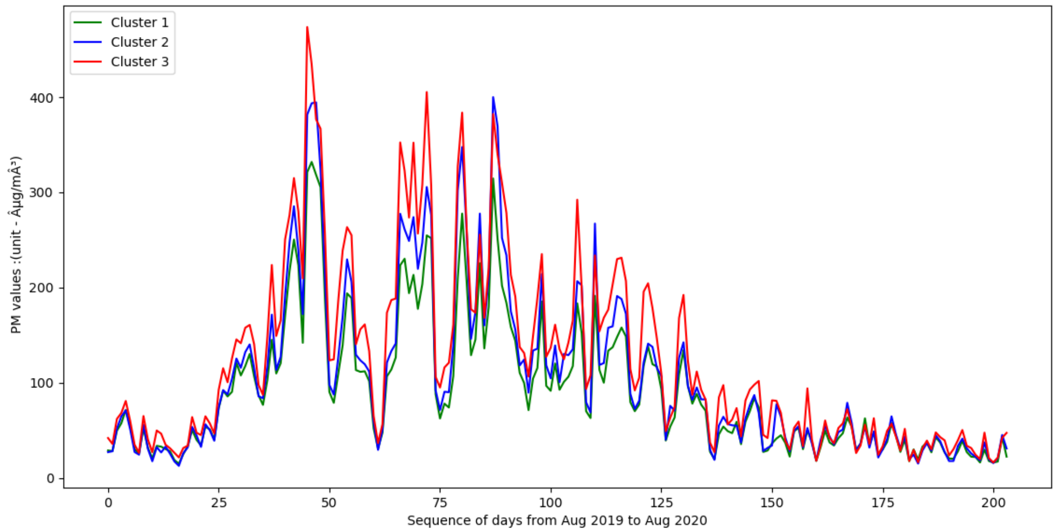


Fig. 9. Average PM values of clusters across the duration of data.

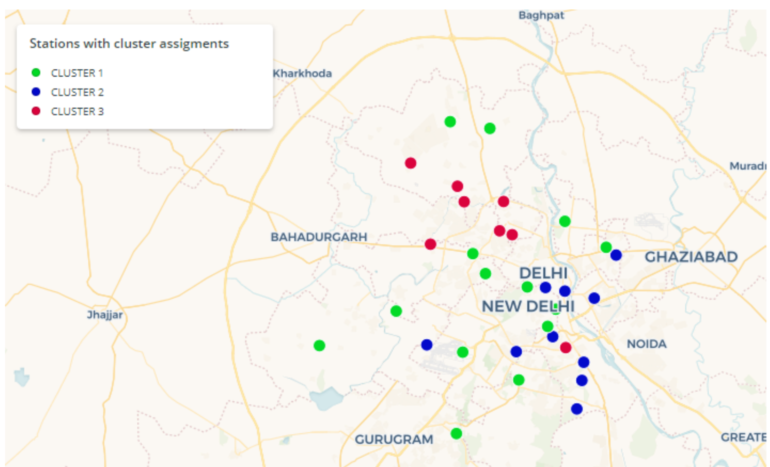


Fig. 10. Location of different sensors belonging to different PM2.5 clusters.

We examine which Google Places categories influence PM 2.5 values most. We model this problem as finding the feature importance of a classification task. The features are the aggregated counts of 10 Google places categories within a 500m buffer around each monitoring station. The classification task is to classify each monitoring station to their respective PM 2.5 cluster, as shown in Figure 10. We compute feature importance using Random forest and also with SVM, and show the results in Figure 11.

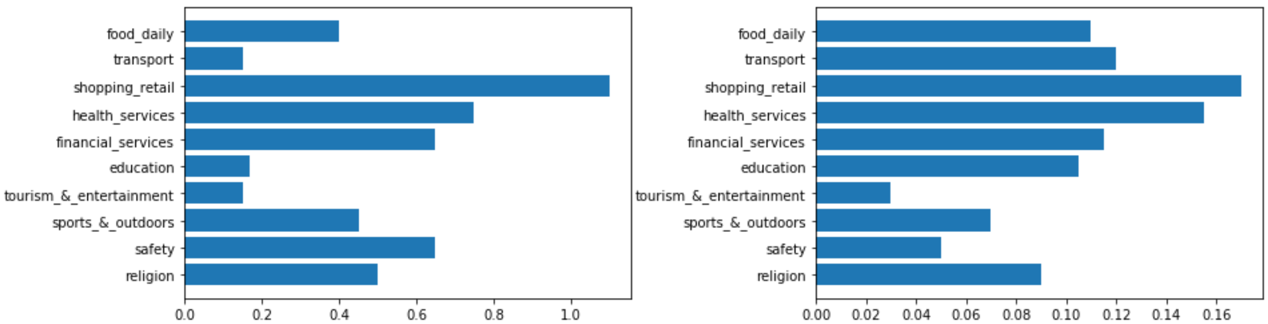


Fig. 11. Feature Importance of different google places features using SVM (left) and Random Forest (right)

3.3 Dynamic Factors: Correlation with road traffic density and meteorological factors

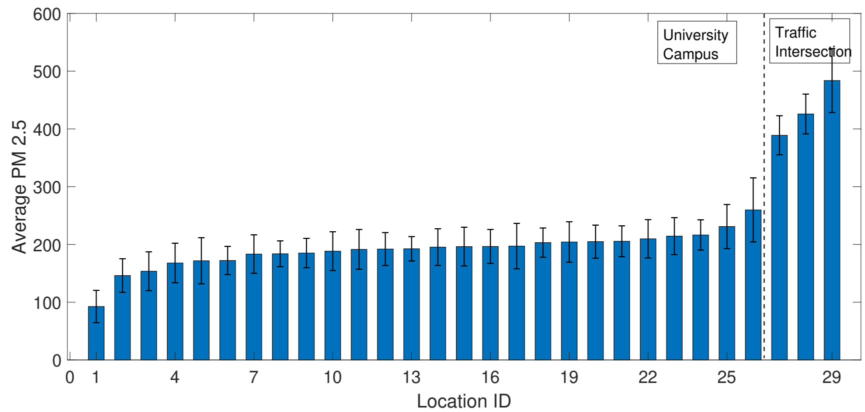
Traffic density and meteorological factors like temperature change rapidly in a city. These factors might also exhibit strong diurnal, weekly and seasonal patterns. We term these factors as dynamic factors. We have taken multiple days of an year as examples to examine their relation with PM in this section.

Figure 12 shows the average PM 2.5 on two days on Nov 2019, as recorded in 29 device deployments (26 instruments in campus and the rightmost 3 instruments at traffic intersections within 3-5 Km radius of the university campus). Figure 13(a) shows locations of 6 sample campus sensors and figure 13(b) shows a sample deployment of our instrument on a pole at a traffic intersection. Campus sensors 11, 24 and 25 record moderate to high PM values, as they see more vehicular traffic near gates. The effect of vehicular emissions is even more prominent for the last three bars in Figure 12(a) and (b), from sensors deployed at busy traffic intersections. While the maximum PM 2.5 value is 250 units on Nov 14 in the campus, on the road it goes to 500 units. On Nov 18 also, the campus (PM 2.5 value 50 units) vs. road (PM 2.5 value 100 units) PM values indicate, that traffic and PM are correlated.

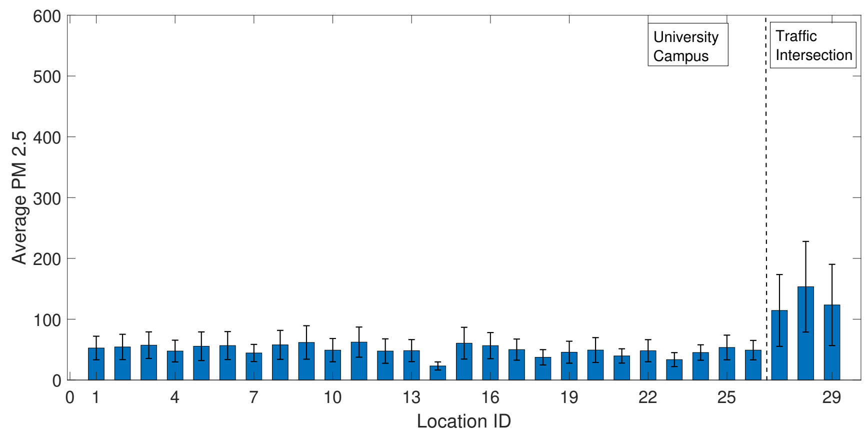
Figure 14(a) shows the PM 2.5 values recorded by the 6 sample sensors in campus (Figure 13(a)) on Nov 18, 2019. There is a strong diurnal trend in the data for all six sensors. Night shows higher and day shows lower values. Thermally induced convection can explain the effect of increased daytime temperatures on PM concentrations. As the ground heats up during the day, gusts and winds increase, leading to increased diffusion of PM. At night (especially in late autumn and winter), decreasing temperatures create a temperature inversion which acts as a cap, inhibiting PM diffusion. Such diurnal trends have been reported in many prior PM characterization papers [22, 33]. Thus PM seems to vary with both traffic density and ambient temperatures, the two dynamic factors we explore in this section.

3.4 When multiple factors co-exist and all affect PM, which factor dominates?

While meteorology and traffic density are both seen to affect PM values, their combined effects is particularly interesting to observe. Figure 15 (a) shows, on the right y-axis, PM levels at a traffic intersection on Nov 18, 2019. The diurnal pattern of high PM values at night and lower values mid-day is consistent as seen with the campus sensors in Figure 14. Interestingly, the left y-axis in Figure 15 (a) shows the median number of vehicles/second, computed over 15 minutes, as measured by the IoT unit's camera and using the vehicle detection Neural Network models in chaotic traffic from [5]. This traffic density value is understandably higher during the day than late night. Thus we have a very interesting observation here: higher temperatures in the day cause PM values to drop, even at very high traffic densities, while the accumulated PM from vehicular emissions shoots up ground level air pollution, when temperature drops at night.



(a) Nov 14, 2019

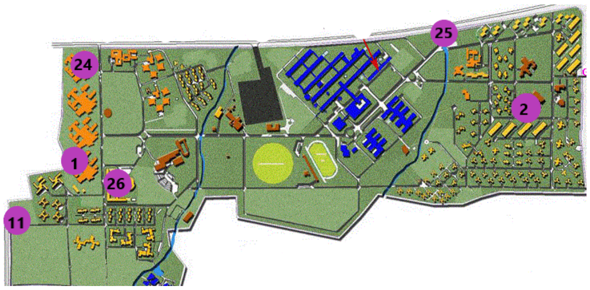


(b) Nov 18, 2019

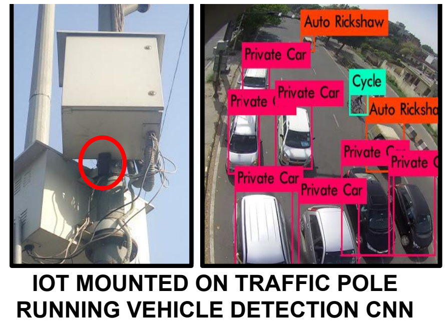
Fig. 12. Average PM 2.5, as recorded by IoT mounted units in campus and traffic intersections

Temperature causes traffic density to not have direct temporal coordination with PM levels, but a delayed correlation. These interesting joint effects cannot be analyzed without a multi-sensor approach, that our IoT device facilitates integrating PM and camera sensor hardware, with automated CNN based accurate and low latency software, for camera data analysis.

In addition to temperature, winds also play a vital role in PM levels. Figure 12 shows PM values on Nov 14 and Nov 18, 2019. The odd-even vehicular policy was enforced in Delhi-NCR between Nov 1 - Nov 15, 2019, i.e. vehicles with odd numbered plates would ply on odd dates and those with even numbered plates would ply on even dates. Thus Nov 14 is supposed to have much less vehicular traffic than on Nov 18, the latter having no odd-even rule in place. However, the PM values are higher on Nov 14 (Figure 12(a)) than on Nov 18 (Figure 12(b)), by a whopping 2-5 times, based on where the sensors are deployed. This deviates from correlation of PM values with traffic density, that we consistently see in the campus sensors deployed near gates vs. sensors which are well within campus, and also in consistent differences seen by the campus vs. the traffic intersection sensors. Figure 14(b) explains this counter-intuitive phenomena. This

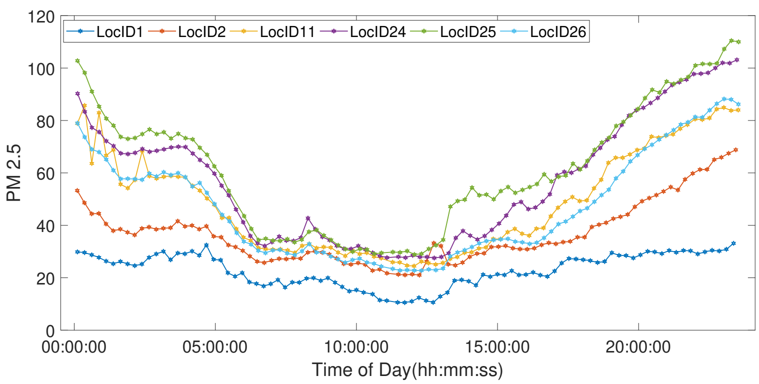


(a) Sample sensor locations in university campus

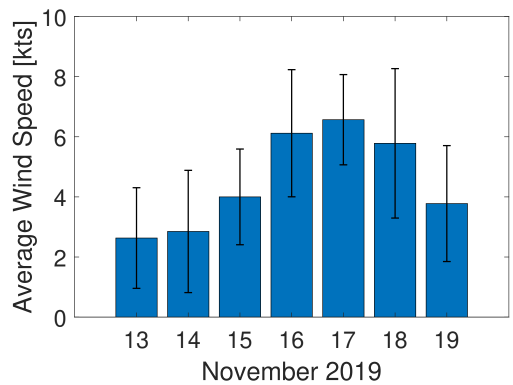


(b) Sensor deployment at traffic intersection

Fig. 13. University campus and traffic intersection deployments of our custom instruments containing PM sensor, camera etc.



(a) Diurnal pattern of PM on Nov 18, 2019



(b) Wind speed in Nov 2019

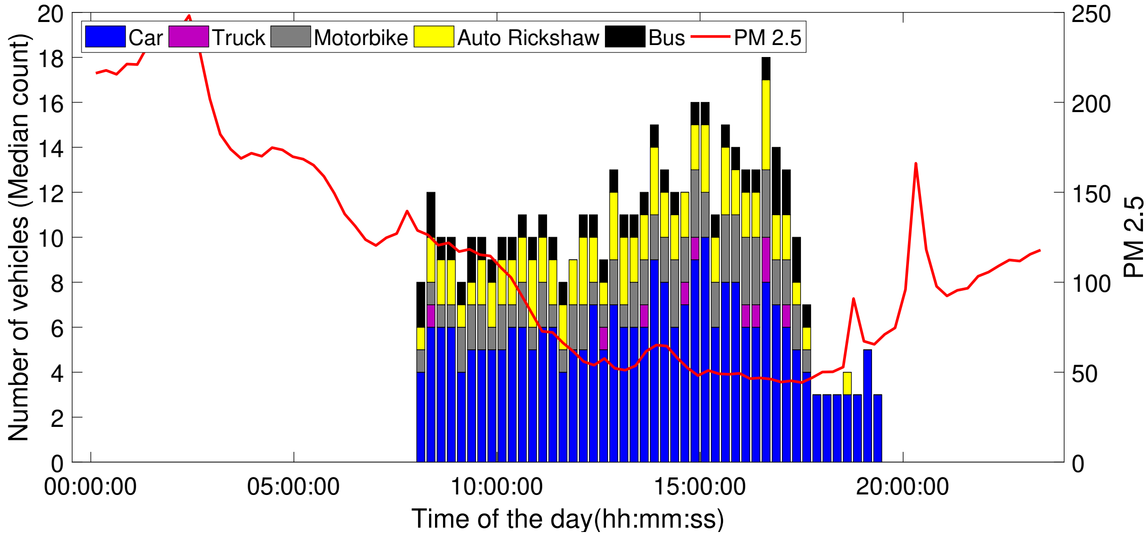
Fig. 14. Meteorological effects on PM values

shows the wind speeds in Delhi-NCR, as measured by the Delhi International Airport authorities, during the period in Nov 2019 under discussion. The wind speeds on Nov 17-18 are significantly higher than that on Nov 13-14. As strong winds help to diffuse PM, it explains the discrepancy in Figure 12, where effect of traffic reduction is overwhelmed by surface winds [47, 49].

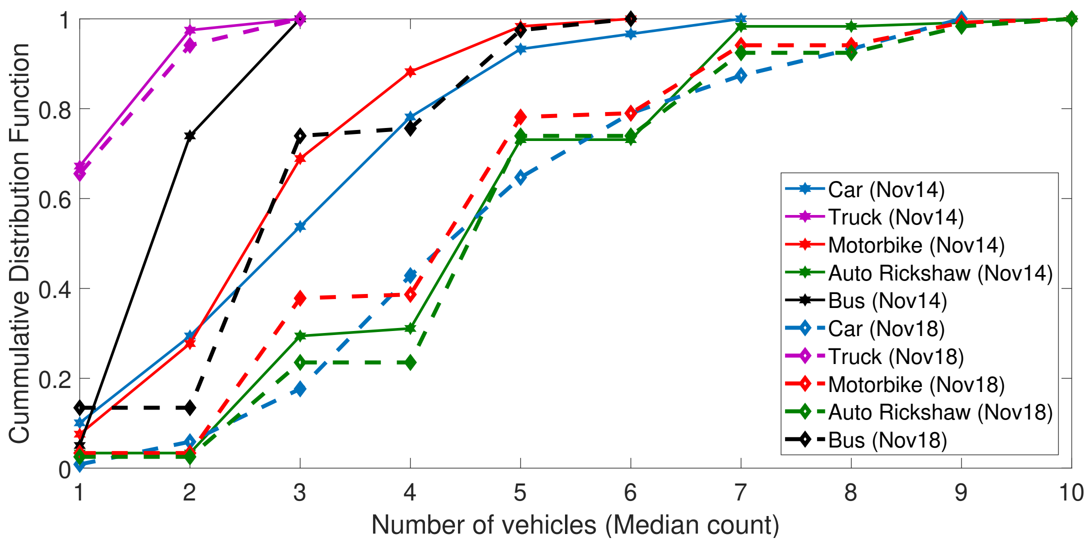
We finally ask the question, did odd-even rule at least reduce vehicular traffic? When multiple factors jointly affect PM values, it is difficult to audit public policies like odd-even rule for their efficacy. A direct measurement of what the policy is trying to achieve (namely reduce number of vehicles), might be a better metric. As seen from Figure 15 (b), median number of vehicles/second, computed over 5 minutes, shows higher values on Nov 18 (without odd-even rule) vs. Nov 14 (with odd-even rule). Thus the policy achieves its primary goal. Whether that in turn reduces PM depends on other factors, which in case of meteorological phenomena is impossible to control by humans. But at least controlling traffic, so that it does not exacerbate high PM values when meteorological conditions are poor, seems reasonable.

4 CONCLUSION AND FUTURE WORK

This paper presents a custom-designed low cost IoT platform and web API based data collection and processing software, to collect PM and auxiliary information about factors affecting PM. We deploy our custom platforms in university



(a) Delayed correlations of PM 2.5 with traffic levels



(b) Vehicle counts with and without odd-even policy

Fig. 15. Combined effects of traffic and meteorological phenomenon on PM values

campus and at traffic intersections, to collect PM, meteorological and traffic density data. We further collect web API based green cover, built-area, residential and commercial area related information. Using these variety of data sources, we analyze the relations between PM and static factors that change slowly in a city (green cover, built area, residential and commercial characteristics), dynamic factors that vary rapidly with diurnal, weekly and seasonal patterns (traffic density and meteorological factors) and combinations of more than one factor. Our main observations are: (a) green cover and PM are negatively correlated, (b) built area and PM are positively correlated, (c) commercial vs. residential areas do not show direct correlations, but there are place categories that can be ranked to correlate with PM levels, (d) temperature and PM are negatively correlated, (e) traffic density and PM are positively correlated, (f) when temperature and traffic are both high during daytime, temperature dominates to make PM lower, but the high traffic density during day shows a delayed correlation causing higher PM values at night and (g) high winds can lower PM values, even if traffic densities are higher, as seen with and without the odd even policy in our deployment.

In future, we will scale up these PM measurements and auxiliary data collection over larger areas of Delhi-NCR and Northern India. We are already deploying these sensors in Delhi public buses in collaboration with Delhi Integrated Multimodal Transit System (DIMTS). We will make all non-proprietary datasets, collected by us, public. We will also do more nuanced and extensive factor correlation analyses, using the gradually increasing datasets and Machine Learning based factor analysis methods [2].

REFERENCES

[1] Ankush Gupta Abhishek Dutta and Andrew Zisserman. [n.d.]. VGG image annotator. https://www.robots.ox.ac.uk/~vgg/software/via/.

[2] Nipun Batra, Hongning Wang, Amarjeet Singh, and Kamin Whitehouse. 2017. Matrix Factorisation for Scalable Energy Breakdown. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, California, USA) *(AAAI'17)*. AAAI Press, 4467–4473.

[3] Matthias Budde, Mathias Busse, and Michael Beigl. 2012. Investigating the use of commodity dust sensors for the embedded measurement of particulate matter. In *2012 Ninth International Conference on Networked Sensing (INSS)*. IEEE, 1–4.

[4] Matthias Budde, Rayan El Masri, Till Riedel, and Michael Beigl. 2013. Enabling low-cost particulate matter measurement for participatory sensing scenarios. In *Proceedings of the 12th international conference on mobile and ubiquitous multimedia*. 1–10.

[5] Mayank Singh Chauhan, Arshdeep Singh, Mansi Khemka, Arneish Prateek, and Rijurekha Sen. 2019. Embedded CNN Based Vehicle Classification and Counting in Non-laned Road Traffic. In *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development (ICTD '19)*.

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. 2016. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *CoRR* abs/1606.00915 (2016). arXiv:1606.00915 http://arxiv.org/abs/1606.00915

[7] Yun Cheng, Xiaoxi He, Zimu Zhou, and Lothar Thiele. 2019. Ict: In-field calibration transfer for air quality sensor deployments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–19.

[8] Yun Cheng, Xiucheng Li, Zhijun Li, Shouxu Jiang, Yilong Li, Ji Jia, and Xiaofan Jiang. 2014. AirCloud: A Cloud-based Air-quality Monitoring System for Everyone. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems (SenSys '14)*.

[9] Central Pollution Control Board (CPCB). 2016. Delhi Odd-Even Data. http://cpcb.nic.in/delhi-odd-even-data/.

[10] Nandini Dasgupta. 2015. Tall Blunder. https://www.downtoearth.org.in/coverage/tall-blunder-22419.

[11] Down To Earth. 2018. Crop burning: Haryana farmers to launch a state-wide protest. https://www.downtoearth.org.in/news/air/crop-burning-haryana-farmers-to-launch-a-state-wide-protest-61889.

[12] Down To Earth. 2018. Crop burning: Why are Punjab farmers defying government ban. https://www.downtoearth.org.in/news/air/crop-burning-why-are-punjab-farmers-defying-government-ban-61869.

[13] Ecotech. 2016. Odd-Even Policy, Delhi, Explained. https://www.ecotech.com/odd-even-policy-delhi-explained.

[14] The Indian Express. 2018. 14,000 of 21,000 trees to be axed for redevelopment of south Delhi colonies: Govt. http://tinyurl.com/ybys6zro.

[15] The Indian Express. 2018. Delhi tree felling: Death by a thousand cuts. https://indianexpress.com/article/cities/delhi/delhi-tree-cutting-forest-cover-pollution-5241664/.

[16] Carnegie Council for Ethics in International Affairs. 2004. Workers' Rights and Pollution Control in Delhi. https://www.carnegiecouncil.org/publications/archive/dialogue/2_11/section_2/4451.

[17] Forbes. 2016. What Did Delhi's Odd-Even Scheme Achieve? Not Much. https://www.forbes.com/sites/meghabahree/2016/01/18/what-did-delhis-odd-even-scheme-achieve-not-much/#3eb123dd15cc.

[18] Jiahao Gao, Zhiwen Hu, Kaigui Bian, Xinyu Mao, and Lingyang Song. 2020. AQ360: UAV-aided air quality monitoring by 360-degree aerial panoramic images in urban areas. *IEEE Internet of Things Journal* 8, 1 (2020), 428–442.

[19] Meiling Gao, Junji Cao, and Edmund Seto. 2015. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2. 5 in Xi'an, China. *Environmental pollution* 199 (2015), 56–65.

[20] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. *CoRR* abs/1703.06870 (2017). arXiv:1703.06870 http://arxiv.org/abs/1703.06870

[21] Deccan Herald. 2016. Delhi's odd-even scheme has no impact: study. https://www.deccanherald.com/content/666902/delhis-odd-even-scheme-has.html.

[22] German Hernandez, Terri-Ann Berry, Shannon L Wallis, and David Poyner1. 2017. Temperature and Humidity Effects on Particulate Matter Concentrations in a Sub-Tropical Climate During Winter. *International Proceedings of Chemical, Biological and Environmental Engineering* 102 (2017).

[23] The Hindu. 2016. Levels of pollutants increased during odd-even. https://www.thehindu.com/news/cities/Delhi/levels-of-pollutants-increased-during-odd-even/article23166208.ece.

[24] Ke Hu, Vijay Sivaraman, Blanca Gallego Luxan, and Ashfaqur Rahman. 2015. Design and evaluation of a metropolitan air pollution sensing system. *IEEE Sensors Journal* 16, 5 (2015), 1448–1459.

[25] Aditya Nigam in Revolutionary Democracy. 2001. The Hindu. https://www.thehindu.com/news/national/other-states/punjabs-burning-problem/article25339426.ece.

[26] Aditya Nigam in Revolutionary Democracy. 2001. Industrial Closures in Delhi. http://www.revolutionarydemocracy.org/rdv7n2/industclos.htm.

[27] Quartz India. 2016. Odd-even won't fix Delhi's air, but it will feed the selfishness of Delhi's car owners. http://tinyurl.com/y2vgxgwg.

[28] Wan Jiao, Gayle Hagler, Ronald Williams, Robert Sharpe, Ryan Brown, Daniel Garver, Robert Judge, Motria Caudill, Joshua Rickard, Michael Davis, et al. 2016. Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmospheric Measurement Techniques* 9, 11 (2016).

[29] Jinchao Li, Lin Chen, Yuwei Xiang, and Ming Xu. 2018. Research on influential factors of PM2.5 within the beijing-tianjin-hebei region in China. *Discrete Dynamics in Nature and Society* 2018 (2018).

[30] Jason Jingshi Li, Boi Faltings, Olga Saukh, David Hasenfratz, and Jan Beutel. 2012. Sensing the air we breathe—the OpenSense Zurich dataset. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

[31] Yansui Liu, Yang Zhou, and Jiaxin Lu. 2020. Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Scientific reports* 10, 1 (2020), 1–11.

[32] The Times of India. 2018. No more trees will be cut for south Delhi projects: Centre. https://timesofindia.indiatimes.com/city/delhi/no-more-trees-will-be-cut-for-south-delhi-projects-centre/articleshow/64786047.cms.

[33] P.D.Hiena, V.T.Bacb, H.C.Thamb, D.D.Nhanb, and L.D.Vinh. 2002. Influence of meteorological conditions on PM2.5 and PM3.5-10 concentrations during the monsoon seasons in Hanoi, Vietnam. *Atmospheric Environment* 36, 21 (2002).

[34] The Pioneer. 2017. Farmers protest Punjab Government's orders. https://www.dailypioneer.com/2017/state-editions/farmers-protest-punjab-governments-orders.html.

[35] First Post. 2016. New study says odd-even scheme in Delhi led to an increase in number of vehicles and hence, emissions. http://tinyurl.com/y2clsmbh.

[36] Aakash C Rai, Prashant Kumar, Francesco Pilla, Andreas N Skouloudis, Silvana Di Sabatino, Carlo Ratti, Ansar Yasar, and David Rickerby. 2017. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Science of The Total Environment* 607 (2017), 691–705.

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR* abs/1505.04597 (2015). arXiv:1505.04597 http://arxiv.org/abs/1505.04597

[38] Scroll.in. 2016. Delhi responds enthusiastically to odd-even policy's first real test. https://scroll.in/article/801356/delhi-responds-enthusiastically-to-odd-even-policys-first-real-test.

[39] Scroll.in. 2018. Opinion: Why felling of thousands of trees for South Delhi redevelopment project must be reviewed. http://tinyurl.com/y6nyeloo.

[40] Amanjeet Singh and Hanan Zaffar. 2017. Why Do Farmers Burn the Crop Residue in Punjab? https://www.newsclick.in/why-do-farmers-burn-crop-residue-punjab.

[41] Reka Srinivas, Abhilash S Panicker, Neha S Parkhi, Sunil K Peshin, and Gufran Beig. 2016. Sensitivity of online coupled model to extreme pollution event over a mega city Delhi. *Atmospheric Pollution Research* 7, 1 (2016), 25–30.

[42] Amos PK Tai, Loretta J Mickley, and Daniel J Jacob. 2010. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. *Atmospheric environment* 44, 32 (2010), 3976–3984.

[43] Hindusthan Times. 2016. Air cleaner this April than last year, says body studying odd-even. https://tinyurl.com/y4uk9u47.

[44] Hindusthan Times. 2016. Chaos, jams rule Delhi roads on return of odd, even scheme. https://www.hindustantimes.com/delhi-news/odd-even-repairs-breakdowns-chaos-jams-ruled-delhi-roads-on-monday/story-e7pemlXHQ4KN2xadjBiJdL.html.

[45] Hindustan Times. 2018. 16,500 trees: A huge price for south Delhi's redevelopment projects. https://tinyurl.com/y73te44m.

[46] Hindustan Times. 2018. One tree cut every hour over last 13 years, says Delhi govt data. https://www.hindustantimes.com/delhi-news/one-tree-cut-every-hour-over-last-13-years-says-delhi-govt-data/story-uJBiGcLemQIOCvIfP7rwpN.html.

[47] Hindusthan Times. 2019. Slow wind, low temp push Delhi air towards emergency zone. https://www.hindustantimes.com/delhi-news/slow-wind-low-temp-push-delhi-air-towards-emergency-zone/story-rMV50V6ygPcVpzIgbH8FXO.html.

[48] Suresh Tiwari, Atul Kumar Srivastava, Deewan Singh Bisht, Pragya Parmita, Manoj K Srivastava, and SD Attri. 2013. Diurnal and seasonal variations of black carbon and PM2. 5 over New Delhi, India: Influence of meteorology. *Atmospheric Research* 125 (2013), 50–62.

[49] India Today. 2019. High surface winds significantly improve Delhi's air quality to poor category. https://www.indiatoday.in/india/story/delhi-air-quality-index-november18-1619938-2019-11-18.

[50] Ravi Yadav, LK Sahu, G Beig, and SNA Jaaffrey. 2016. Role of long-range transport and local meteorology in seasonal variation of surface ozone and its precursors at an urban site in India. *Atmospheric Research* 176 (2016), 96–107.

[51] Yuzhe Yang, Zhiwen Hu, Kaigui Bian, and Lingyang Song. 2019. Imgsensingnet: Uav vision guided aerial-ground air quality sensing system. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 1207–1215.

[52] Tongshu Zheng, Michael H. Bergin, Karoline K. Johnson, Sachchida N. Tripathi, Shilpa Shirodkar, Matthew S. Landis, Ronak Sutaria, and David E. Carlson. 2018. Field evaluation of low-cost particulate matter sensors in high and low concentration environments. *Atmospheric Measurement Techniques* (2018).